TEKNIK SMOTE PADA KLASIFIKASI SENTIMEN PRODUK KECANTIKAN VIRAL DI AKUN X @OHMY_BEAUTYBANK MENGGUNAKAN ALGORITMA C4.5

Anisa Faozyah, Anisa Lutfiyani
Universitas Ma'arif Nahdlatul Ulama Kebumen
Jl. Kutoarjo No.Km.05, Wonoboyo, Jatisari, Kec. Kebumen, Kabupaten Kebumen, Jawa Tengah 54317
anisafaozyah63@gmail.com

ABSTRAK

Pesatnya pertukaran informasi dan opini konsumen mengenai produk kecantikan di media sosial telah menjadi fenomena yang berpengaruh besar terhadap keputusan pembelian di era digital saat ini. Penelitian ini mengamati aktivitas tersebut pada akun X (Twitter) *@ohmy_beautybank*, di mana ulasan pengguna mengandung informasi berharga namun terkendala oleh masalah ketidakseimbangan data (*imbalanced data*) yang mencolok antara sentiment positif dan negatif. Ketimpangan ini beresiko menyebabkan model klasifikasi menjadi kurang objektif dalam memberikan prediksi. Penelitian ini bertujuan untuk meningkatkan performa klasifikasi sentiment agar tetap akurat meskipun terdapat sebaran data yang tidak merata. Untuk mencapai tujuan tersebut, diterapkan algoritma C4.5 yang diintegrasikan dengan Teknik *Synthetic Minority Oversampling Technique* (SMOTE) untuk menyeimbangkan distribusi kelas, serta penggunaan fitur N-Grams (*Bigram)* untuk menangkap konteks makna pada ulasan. Hasil eksperimen menunjukkan bahwa integrasi metode tersebut efektif meningkatkan performa model dengancapaian akurasu sebesar 81.20%. dapat disimpulkan bahwa penggunaan SMOTE dan N-Grams berhasil mengatasi kelemahan algoritma C4.5 dalam menangani data tidak seimbang, sehingga model ini layak digunakan sebagai alat analisis opini pasar yang lebih akurat dan objektif.

Kata kunci : *Analisis Sentimen, Algoritma C4.5, SMOTE, N-Grams, Twitter.*

1. PENDAHULUAN

Pemanfaatan *Electronic Word of Mouth* (e-WOM) pada media sosial X (Twitter), khususnya akun *@ohmy_beautybank*, telah bertransformasi menjadi rujukan kursial bagi konsumen dalam memvalidasi kualitas produk kecantikan sebelum melakukan pembelian. Informasi yang tersebar dalam ulasan pengguna merupakan asset data penting,baik bagi konsumen sebagai dasar pertimbangan maupun bagi *owner brand* untuk memahami presepsi pasar [1]. Namun, karakteristik data teks yang massif dan tidak terstruktut ditandai dengan dominasi Bahasa *slang* serta gangguan (*noise)* berupa interaksi *spam* menjadi kendala utama jika pemantauan dilakukan secara manual. Kondisi inilah yang memicu *urgensi* diterapkannya klasifikasi sentimen otomatis agar pemilahan opini publik dapat berjalan lebih objektif dan efisien.

Persoalan teknis yang muncul dalam pengolahan data pada akun tersebut adalah adanya ketimpangan kelas yang signifikan (*imbalanced data*). Dalam *dataset @ohmy_beautybank*, jumlah ulasan berlabel positif ditemukan jauh lebih mendominasi dibandingkan ulasan negatif. Ketidakseimbangan ini menjadi ancaman serius bagi performa model klasifikasi karena algoritma akan cemderung memiliki bias tinggi terhadap kelas mayoritas. Meskipun algoritma C4.5 dikenal mumpuni dalam menghasilkan pohon keputusan yang transparan [2], kinerjanya tetap akan terdegradasi secara drastis jika dipaksa bekerja pada sebaran data yang tidak merata.

Berdasarkan permasalahan tersebut, rencana penyelesaian masalah dalam penelitian ini difokuskan pada integrasi teknik *Synthetic Minority Oversampling Technique* (SMOTE) untuk menstabilkan distribusi data. Tujuan utama dari penelitian ini adalah mengoptimalkan kemampuan algoritma C4.5 dalam mengenali pol akelas minoritas dengan membangkitkan sampel sintetis melalui SMOTE. Selain itu, ekstraksi fitur juga dioerkuat menggunakan N-Grams (*Bigram*) guna menangkap konteks ulasan yang lebih spesifik. Pendekatan ini diharapkan mampu meningkatkan akurasi dan reliabilitas model klasifikasi sentiment, sehingga hasil yang diperoleh tetap valid meskipun dihadapkan pada ketimpangan data yang mencolok [3] .

2. TINJAUAN PUSTAKA

2.1. *Electronic Word of Mouth* (e-WOM)

*Electronic Word of Mouth (*e-WOM*)* merupakan segala pernyataan positif maupun negatif yang dibuat oleh konsumen mengenai suatu produk atau Perusahaan yang tersedia bagi banyak orang melalui media internet. Ulasan produk di media social bertindak sebagai cermin opini public yang mempengaruhi persepsi dan keputusan pembelian konsumen lain [1]. Dalam penelitin ini, e-WOM difokuskan pada akun X (twitter) *@ohmy_beautybank.*

2.2. Analisis Sentimen

Analisis sentimen atau *opinion mining* Adalah proses komputasi data tekstual secara otomatis guna memperoleh informasi polaritas opini. Efektifitas analisis sentimen pada media sosial sangat bergantung pada kualitas *preprocessing* data untuk meminimalisir data berupa Bahasa gaul atau simbol. Proses ini bertujuan mengklasifikasikan teks ke dalam kategori positif atau negatif [4].

2.3. Algoritma C4.5

Algoritma C4.5 model klasifikasi data mining yang bekerja dengan mentransformasi data menjadi pohon keputusan *(decision tree).* Menurut Dualingga [5] algoritma ini sangat efektif untuk klasifikasi teks di media sosial karena mampu menangani data numerik maupun kategorikal. Proses pembuatan pohon keputusan ini didasarkan pada pemilihan atribut dengan nilai *Gain* tertinggi sebagai akar *(root node)*[5].

Untuk menentukan atribut mana yang menjadi akar, dilakukan perhitungan nilai *Entropy* untuk mengukur Tingkat ketidakpastian dalam *dataset*. Sebagaimana dijelaskan dalam penelitian Marcillia dkk.[6] , rumus dasar yang digunakan Adalah sebagai berikut :

$$Entropy\ (S) = \sum_{i=1}^{n} -p_i \log_2 p_i$$

Keterangan : *S* adalah himpunan sampel data, dan p_i Adalah proporsi atau peluang kelas ke -*I* dalam *dataset* tersebut. Setelah nilai *Entropy* didapat, mesin akan menghitung nilai *Gain* untuk tiap atribut. Strktur pohon keputusan ini secara teoritis terdiri dari akar, cabang, dan daun yang berfungsi sebagai label klasifikasi.

2.4. *Synthetic Minority Oversampling Technique* (SMOTE)

Teknik SMOTE bekerja pada ruang fitur dengan cara menciptakan sampel sintesis yang menghubungkan tetangga terdekat pada kelas minoritas. Berbeda dengan *oversampling* tradisional, SMOTE melakukan interpolasi antara sampel asli sehingga menghasilkan selebaran data yang lebih beragam. Langkah ini kursial menjadi bias terhadap kelas mayoritas [3]. Selain itu penyeimbangan data melalui SMOTE terbukti mampu mengoptimalkan performa klasifikasi secara signifikan pada *dataset* yang jomplang [7].

2.5. Penelitian Terdahulu

Penelitian ini merujuk pada beberapa studi relevan guna memposisikan kebaruan riset yang dilakukan. Studi awal menekankan bahwa efektifitas analysis sentimen sangat bergantung pada kualitas pengolahan data tekstual [4]. Dalam penerapannya, sebelumnya telah berhasil membuktikan keandalan algoritma C4.5 dalam mengklasifikasi opini punlik pada platform media sosial [5][2]. Namun, kendala utama yang sering muncul dalam *dataset* media sosial adalah ketimpangan jumlah data, di manA penggunaan teknik SMOTE sangat efektif menyeimbangkan distribusi kelas tersebut [8].

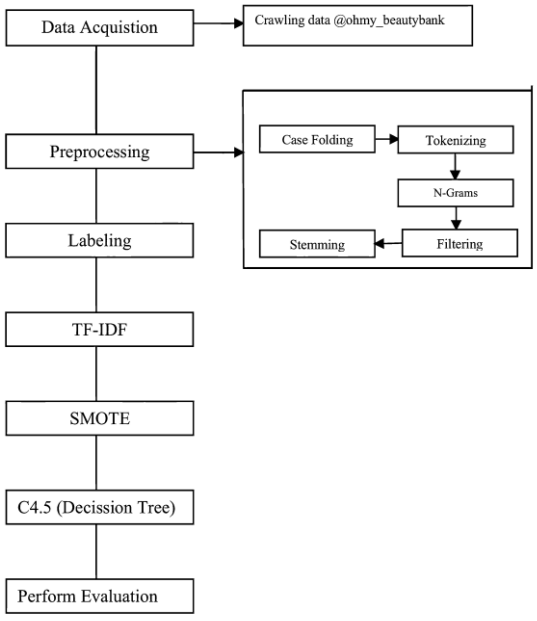
Selain itu keunggulan algoritma berbasis pohon keputusan ini juga didukung oleh temuan yang menunjukkan bahwa c4.5 memilik performa yang lebih baikdalam hal interpretasi aturankeputusan jika dibandingkan dengan metode Naïve Bayes [9]. Hal ini selaras dengan hasil penelitian lain yang menyatakan bahwa implementasi SMOTE terbukti mampu mengatasi masalah imbalanced data dan secara signifikan meningkatkan hasil evaluasi pada analisis sentimen yang memiliki sebaran kelas tidak merata [10]. Guna mengoptimalkan ekstraksi fitur teks, penggunaan TF-IDF yang dikombinasikan dengan N-Grams mampu meningkatkan kinerja klasifikasi karena dapat menangkap konteks frasa yang lebih spesifik dibandingkan kata Tunggal [11]. Berdasarkan landasan tersebut, posisi penelitian ini Adalah menginteregasikan Teknik SMOTE, fitur N-Grams, dan algoritma C4.5 secara bersamaan. Fokus utama riset ini diarahkan pada domain ulasan produk kecantikan pada akun X(Twitter) *@ohmy_beautybank*, yang memiliki karakteristik Bahasa unik dan dinamika opini yang sangat cepat dibandingkan objek penelitian pada studi – studi sebelumnya.

3. METODE PENELITIAN

Metode penelitian ini disusun untuk memberikan Gambaran sistematis mengenai tahapan pengolahan ulasan dari akunX (Twitter) @ohmy_beautybank. Proses ini dimulai dari akusisi data hingga pengujian performa algoritma C4.5 yang telah diintegrasikan dengan SMOTE.

3.1. Tahapan Penelitian

Alur kerja penelitian ini dibagi menjadi beberapa tahap utama yang saling berkaitan. Secara visual tersebut dapat dilihat pada Gambar 1 berikut:



Gambar 1. Flowchart Penelitian

3.2. Akusisi Data (Data Acquistion)

Data ulasan diperoleh melalui proses crawling pada akunX @ohmy_beautybank menggunakan kata kunci yang relevan dengan ulasan produk kecantikan. Melalui proses ini, peneliti berhasil mengumpulkan sebanyak 600 data ulasan. Data yang terkumpul

kemudian disimpan dalam format .csv sebagai *dataset* mentah yang akan diproses pada tahap berikutnya

Tabel 1. *Dataset*

No	Tweet
1.	Aku pakai skin 1004 whiting ampule dan cream nya juga. Udah jalan 3 tahun dan muka ku cerahan gak kusem
2.	hanasui vit c pleaseee!!!! aku udah beli 2 kali gongg banget cerahnyaa
3.	aku pake tinted sunscreen lucienne, baguss banget soalnya finishnya glowing
3.	BIG THANKS to Judydoll karna udah nyiptain lip product yang bisa bener-bener NUTUPIN my TWO TONES LIPS
4.	hanasui vit c mantulll nder

Berdasarkan Tabel 1, terlihat bahwa data ulasan diperoleh dari akun *@ohmy_beautybank* memiliki karakteristik teks yang tidak terstruktur, penggunaan *slang*, serta penggunaan simbol atau emoji yang intensif. Sebaran awal data menunjukkan dominasi sentimen positif yang sangat signifikan dibandingkan sentimen negatif. Hal ini menunjukkan adanya masalah *imbalanced data* yang dapat menyebabkan model klasifikasi memiliki kecenderungan bias terhadap kelas mayoritas jika tdak ditangani dengan Teknik penyeimbang data.

3.3. *Preprocessing*

Pada tahap ini, data mentah hasil *crawling* ulasan produk kecantikan dari akun X (*Twitter*) *@ohmy_beautybank* diproses agar menjadi data terstruktur. Serangkaian tahapan *preprocessing* dilakukan untuk menghilangkan *noise* dan menyeragamkan format teks. Mengingat ulasan media sosial banyak menggunakan Bahasa gaul dan singkatan, langkah – langkah yang diambil Adalah sebagai berikut :

3.4. Tokenizing

Langkah awal untuk memecah kalimat ulasan menjadu potongan kata tunggal (token). Hal ini bertujuan agar setiap kata dapat diidentifikasi secara mandiri sebelim dibentuk jadi fitur frasa.

3.5. N – Grams

Penerapan *bigram* dilakukan untuk membenyuk pasangan dua kata yang berurutan. Langkah ini sangat penting untuk menangkap konteks yang sering muncuk pada ulasan produk, seperti frasa "tidak cocok" atau "bikin kering", yang maknanya akan hilang jika hanya menggunakan kata Tunggal (*unigram).*

3.3.2. Transform Cases

Seluruh kata yang telah di tokenize diubah menjadi huruf kecil (*lowercase).* Tujuannya agar sistem tidak membaca dua kata yang sama sebagi fitur yang berbeda hanya karena perbedaan penggunaan huruf kapital.

3.6. Filtering (Stopwords & Tokens)

Kata – kata yang tidak memiliki bobot sentimen seperti: "dan", "yang", "di" dihapus menggunakan daftar *stopword* Bahasa Indonesia. Selain itu, dilakukan filter berdasarkan Panjang karakter (*length)* untuk membuang token di bawah 3 huruf yang biasanya berupa *typo* atau simbol sisa.

3.7. Stemming

Tahap akhir *preprocessing* Adalah mengembalikan kata ke bentuk dasarnya guna meredukasi dimensi fitur. Dalam praktiknya, beberapa Bahasa gaul (*slang)* tetap pada bentuk aslinya karena tidak terdeteksi oleh kamus pemotong akar kata standar, namun fitur tersebut tetap dipertahankan karena mimiliki informasi sentimen yang kuat. Sebagai ilustaris data pada setiap tahapan, Tabel 2 menyajikan informasi teks yang terjadi.

Tabel 2. Contoh Transformasi Data pada Tahap Preprocessing

Tahapan	Hasil transformasi teks
Data mentah	"Agak kureng sih. Ga ngehidrasi dan buat oiy skin jadi makin kusam"
Tokenizing	[Agak], [Kureng],[sih],[Ga],[ngehidrasi],[dan],[bikin],[oily],[skin],[jadi],[makin],[kusam].
N-Grams	[Agak_kureng],[kureng_sih],[sih_Ga],[Ga_ngehidrasi], [oily_skin],[skin_jadi],[jadi_makin].[makin_kusam]
Transform cases	[agak_kureng], [kureng_sih], [sih_ga], [ga_ngehidrasi], [oily_skin], [skin_jadi], [jadi_makin], [makin_kusam].
Filtering	[agak_kureng], [ngehidrasi], [oily_skin], [skin_jadi], [kusam].
Stemming	[agak_kureng], [hidrasi], [oily_skin], [skin_jadi], [kusam].

Proses transformasi data pada Tabel 2 menunjukkan bahwa langkah *filtering* dan *stemming* berhasil meredukasi *noise* tanpa menghilangkan esensi makna dari ulasan tersebut. Penggunaan fitur N-Grams (*Bigram*) pada tahap ini sangat kursial. sebagai contoh, frasa 'agak_kureng' tetap dipertahankan sebagai satu kesatuan fitur. Jika menggunakan *unigram,* kata 'agak' akan terpisah dari 'kureng', sehingga informasi sentimen negatif terkandunf di dalamnya menjadi melemah atau hilang sama sekali bagi algoritma C4.5.

3.4. TF-IDF

Setelah data selesai dibersihkan, langkah berikutnya adalah mengubah teks tersebut menjadi nilai numerik menggunakan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Metode ini bekerja dengan menghitung seberapa sering suatu frasa muncul dalam sebuah ulasan dibandingkan dengan distribusinya di seluruh *dataset*.

Penggunaan TF-IDF sangat kursial karena mampu memberikan bobot lebih tinggi pada frasa – frasa unik yang meiliki informasi sentimen kuat. Proses pembobotan ini juga mencakup fitur N-Grams (*Bigram*) yang sudah dibentuk sebekumnya. Sebagai contoh, frasa seperti "bikin_k" atay "skin_jadi" akan mendapatkan bobot numerik tertentu yang nantinya menjadi dasar bagi algoritma C4.5 dalam menyusun struktur pohon keputusan.

3.5. SMOTE

Ditemukan adanya ketidak seimbangan data yang cukup signifikan antara ulasan bersentimen positif dan negatif pada *dataset* yang terkumpul. Untuk menghindari hasil klasifikasi yang bias pada kelas mayoritas, diterapkan teknik SMOTE (*Synthetic Minority Oversampling Technique*). Teknik ini bekerja dengan menciptakan data sintetis baru untuk kelas minoritas (sentimen negatif) berdasarkan kemiripan pola data numerik hasil TF-IDF. Dengan distribusi kelas yang lebih seimbang, algoritma C4.5 dapat mempelajari ciri – ciri ulasan negatif secara mendalam, sehingga model yang dihasilkan tidak hanya akurat pada ulasan positif saja.

3.6. Klasifikasi Algoritma C4.5 dan Evaluasi

Proses klasifikasi dilakukan menggunakan algoritma C4.5 untuk menghasilkan model berupa pohon keputusan yang mudah dipahami logikanya. Untuk memastikan perfotma model yang dibuat, pengujian dilakukan dengan skema *10-Fold Cross Validation*. *Dataset* dibagi menjadi 10 bagian, Di mana proses pelatihan dan pengujian dilakukan sebanyak 10 kali secara bergantian. Hasil akhir diukur menggunakan Confusion Matrix untuk mendapatkan nilai *Accuracy*, *Precision*, dan *Recall*. Evaluasi ini menjadi bukti sejauh mana integrasi SMOTE, N-Grams, dan C4.5 efektif dalam memetakan opini publik pada akun X (Twitter) *@ohmy_beautybank.*

4. HASIL DAN PEMBAHASAN

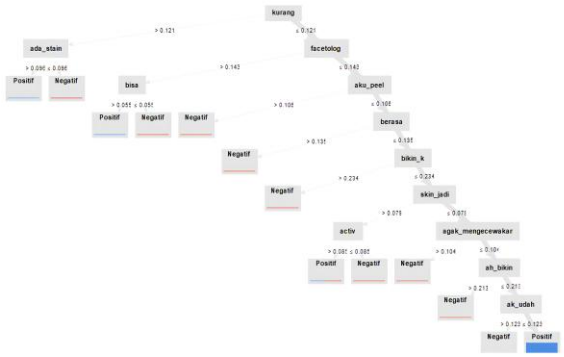
4.1. Analisis Model Klasifikasi C4.5 dan SMOTE

Pada tahap ini, *dataset* ulasan produk kecanikan dari akun X (Twitter) *@ohmy_beautybank* diolah menggunakan algoritma C4.5. mengingat adanya ketimpangan jumlah data (*imbalanced data*) anatara ulasan positif dan negatif \, penelitian ini mengintegrasiikan Teknik SMOTE (*Synthetic Minority Oversampling Technique*) untuk mrnyrimbangkan distribusi kelas. Model yang terbentuk kemudian divisualisasikan dalam bentuk pohon keputusan (*Decission Tree*) untuk melihat logika klasifikasinya.

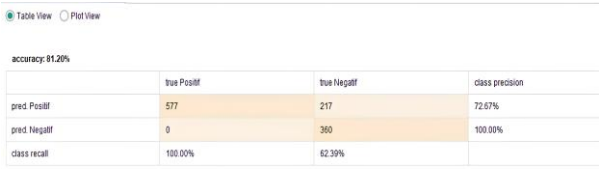
Berdasarkan Gambar 2, terlihat bahwa atribut "kurang" menjadi *Root Node* atau akar utama dalam pembentukan utama dalam pembentukaan pohon keputusan. Menariknya, terdapat beberapa istilah yang mengalami perubahan bentuk akibat proses *stemming* dan *tokenizing*, seperti "facetolog" yang merujuk pada brand Facetology, serta fitur *bigram* seperti "ada_stain" yang menggambarkan konteks ketahanan warna produk. Munculnya kata – kata spesifik brand dan bahasa non-formal seperti "kureng" sebagai node penentu menunjukkan bahwa model mampu menangkap pola komunnikasi unik dan diksi yang sering digunakan audiens di akun *@ohmy_beautybank.*

4.2. Evaluasi Performa Model

Evaluasi terhadap model dilakukan dengan skema 10-Fold Cross Validation. Metode ini dipilih untuk menjamin bahwa nilai akurasu yang diperoleh merupakan representasi dari performa model pada berbagai variasi data, bukan sekedar kebetulan pada satu kali pengujian. Seluruh proses pengujian dijalankan melalui aplikasi RapidMiner, dan hasilnya dirangkum pada Gambar 3.



Gambar 2. Visualisasi Decission Tree Algoritma C4.5



Gambar 3. Performa Klasifikasi Algoritma C4.5

4.3. Pembahasan

Capaian akurasi 81,20% ini membuktikan bahwa kombinasi algoritma C4.5 dan teknik SMOTE bekerja dengan efektif pada *dataset @ohmy_beautybank.* Masalah utama dalam *dataset* ini adalah *class imbalance*,di mana ulasan negatif jumlahnya jauh lebih sedikit disbanding ulasan positif. Tanpa bantuan SMOTE, algoritma cenderung hanya mempelajari kelas mayoritas dan mengabaikan kelas minoritas. Namun, dengan penerapan SMOTE pola – pola ulasan negatif berhasil diperkuat sehingga pohon keputusan dapat mengenali ciri – ciri sentimen negatif secara lebih spesifik melalui aturan – aturan yang terbentuk. Faktor lain yang mendukung performa model adalah penggunaan fitur N-Grams (*Bigram*). Fitur ini memungkinkan sistem menangkap konteks kalimat yang lebih luas dibandingkan hanya menggunakan kata Tunggal (*unigram*). Sebagai contoh "agak kureng" atau "kurang lembab" memberikan informasi

sentimen yang lebih valid bagi algoritma untuk menentukan kelas negatif.

Adapun selisih error sebesar 18.80% kemungkinan besar dipicu oleh ulasan yang mengandung sarkasme atau istilah *slang* yang sangat baru, yang mana masih menjadi tantangan dalam pengolahan Bahasa alami secara otomatis. Meski begitu, secara keseluruhan model ini sudah sangat reliabel untuk memetakan opini publik terhadap produk kecantikan.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa penggabungan algoritma C4.5, SMOTE, dan fitur N-Grams terbukti efektif dalam klasifikasi sentimen dengan akurasi 81.20%. model menunjukkan performa yang sangat kuat dalam menjaring ulasan positif (*Recall* 100%) dan akurat dalam memprediksI sentimen negatif (Precision 100%). Peranan SMOTE sangat kursial dalam menyeimbangkan distribusi data, sementara fitur *Bigram* membantu sistem memahami konteks ulasan non-formal dengan lebih baik. Sebagai saran, penelitian selanjutnya dapat menambahkan kamus normalisasi Bahasa gaul (*slang*) yang lebih komprehensif atau mencoba implementasi algoritma ensemble learning untuk meningkatkan stabilitas prediksi pada *dataset* dengan skala yang lebih besar.

DAFTAR PUSTAKA

[1] K. Jindal and R. Aron, "A Hybrid Machine Learning Approach for Sentiment Analysis of Beauty Products Reviews," *Journal of Information Systems and Telecommunication*, vol. 10, no. 37, pp. 1–10, Dec. 2022, doi: 10.52547/jist.15586.10.37.1.

[2] F. Fersellia, E. Utami, and A. Yaqin, "Sentiment Analysis of Shopee Food Application User Satisfaction Using the C4.5 Decision Tree Method," *Sinkron*, vol. 8, no. 3, pp. 1554–1563, Jul. 2023, doi: 10.33395/sinkron.v8i3.12531.

[3] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiasari, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 173–184, 2021.

[4] O. N. Julianti, N. Suarna, and W. Prihartono, "PENERAPAN NATURAL LANGUAGE PROCESSING PADA ANALISIS SENTIMEN JUDI ONLINE DI MEDIA SOSIAL TWITTER," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 3, 2024.

[5] M. I. Dualingga, E. Budianita, M. Fikry, M. Affandes, and F. Yanto, "Klasifikasi Komentar Terhadap Vaksin Covid-19 Menggunakan Algoritma C4.5 Pada Media Sosial Instagram," in *Seminar Nasional*, 2021, pp. 2579–5406.

[6] M. Marcillia *et al.*, "Analisis Sentimen pada Ulasan Penyedia Layanan Menggunakan Algoritma C4.5," *Journal of Big Data Analytic and Artificial Intelligence*, vol. 6, no. 2, pp. 1–5, 2023.

[7] W. Rahayu, D. Jollyta, A. Hajjah, Y. Nora Marlim, and Y. Desnelita, "Synthetic Minority Oversampling Technique (SMOTE) for Boosting the Accuracy of C4.5 Algorithm Model," *Journal of Artificial Intelligence and Engineering Applications*, vol. 3, no. 3, pp. 2808–4519, 2024, [Online]. Available: https://ioinformatic.org/

[8] N. S. Sediatmoko, Y. Nataliani, and I. Suryady, "Suryady (Sentiment Analysis of Customer Review Using Classification Algorithms and SMOTE for Handling Imbalanced Class)," 2024.

[9] Rovidatul, Y. Yunus, and G. W. Nurcahyo, "Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 193–199, Apr. 2023, doi: 10.37859/coscitech.v4i1.4755.

[10] Erry Maricha Oky Nur Haryanto, Adhien Kenya Anima Estethika, and Rahmad Arif Setiawan, "IMPLEMENTASI SMOTE UNTUK MENGATASI IMBALANCED DATA PADA SENTIMEN ANALISIS SENTIMEN HOTEL DI NUSA TENGGARA BARAT DENGAN MENGGUNAKAN ALGORITMA SVM," *Jurnal Interaksi Interaktif*, vol. 7, no. 1, pp. 16–20, Jan. 2022.

[11] S. Mutmainnah, T. Ansyor Lorosae, and S. Ramadhan, "Model Text Embedding dan TF-IDF+Ngram untuk Meningkatkan Kinerja Algoritma Binary Classifier pada Klasifikasi SMS Palsu," *JURNAL SISTEM INFORMASI TGD*, vol. 4, no. 1, pp. 55–64, 2025, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsi