

KOMPARASI SVM-LR BERBASIS INDOBERT DAN SMOTE-TOMEK PADA DETEKSI BUZZER

Muhamad Hasan Fahreza¹⁾, Nicolas Zefanya Sipayung²⁾, Indra Wahyu Santoso³⁾, Susi Susilowati⁴⁾

^{1,2,3,4)} Informatika Universitas Bina Sarana Informatika

email : 15230766@bsi.ac.id¹⁾, 15230912@bsi.ac.id²⁾, 15230794@bsi.ac.id³⁾, susi.sss@bsi.ac.id⁴⁾

Abstraksi

Dominasi aktivitas *buzzer* politik di TikTok mengancam orisinalitas aspirasi publik melalui manipulasi opini, yang diperburuk oleh ketimpangan distribusi data (*imbalanced data*) sehingga menyulitkan deteksi akurat. Penelitian ini bertujuan mengomparasikan performa *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) berbasis fitur IndoBERT yang dioptimasi dengan *hybrid resampling* SMOTE-Tomek untuk menangani ketidakseimbangan kelas ekstrem. Studi kasus dilakukan pada 1.502 komentar terkait isu penculikan mahasiswa. Hasil pengujian menunjukkan integrasi IndoBERT dan SMOTE-Tomek efektif menghasilkan *Area Under Curve* (AUC) $>0,90$ pada kedua model. Meskipun SVM unggul marginal dalam sensitivitas (AUC 0,9193), *Logistic Regression* terbukti lebih superior dalam menjaga validitas pengguna organik dengan spesifisitas 79,2% dan F1-Score kelas minoritas 0,72. Penelitian ini merekomendasikan LR sebagai model terbaik karena karakteristik probabilistiknya lebih adaptif meminimalisir *False Positive*, sehingga lebih akurat dalam memisahkan narasi organik dari propaganda *buzzer*.

Kata Kunci :

Deteksi Buzzer, IndoBERT, SMOTE-Tomek, *Imbalanced Data*, TikTok.

Abstract

The dominance of political buzzers on TikTok threatens the integrity of public discourse through opinion manipulation, a problem compounded by imbalanced data distributions that hinder accurate detection. This study compares the performance of Support Vector Machine (SVM) and Logistic Regression (LR) utilizing IndoBERT features, optimized with the SMOTE-Tomek hybrid resampling technique to address extreme class imbalance. A case study was conducted on 1,502 comments regarding student kidnapping incidents. Results indicate that integrating IndoBERT with SMOTE-Tomek effectively yields an Area Under Curve (AUC) >0.90 for both models. While SVM shows marginal superiority in sensitivity (AUC 0.9193), Logistic Regression proves superior in preserving organic user validity, achieving a specificity of 79.2% and a minority class F1-Score of 0.72. Consequently, this study recommends LR as the optimal model; its probabilistic nature effectively minimizes False Positives, offering greater accuracy in distinguishing organic narratives from buzzer propaganda.

Keywords :

Buzzer Detection, IndoBERT, SMOTE-Tomek, *Imbalanced Data*, TikTok.

Pendahuluan

Pergeseran diskursus politik dari media berbasis teks seperti X (Twitter) ke video singkat TikTok mengubah lanskap penyebaran informasi secara signifikan. Jika pengguna X relatif terbiasa dengan literasi argumentatif, demografi TikTok didominasi oleh konsumen konten visual. Fenomena ini memvalidasi *Hypodermic Needle Theory*, di mana audiens platform visual cenderung pasif dan rentan dipengaruhi pesan emosional tanpa filter kritis yang kuat[1]. Persoalan utamanya terletak pada agresivitas algoritma *For You Page* (FYP). Data mengindikasikan bahwa algoritma ini memprioritaskan konten emosional dan viral mendominasi hingga 35% total konten politik serta memmarginalkan pesan programatik berbasis data[2].

Kerentanan ekosistem digital ini diperburuk oleh keberadaan *buzzer* (*cyber troops*). Tidak main-main, pemerintah tercatat pernah mengalokasikan dana

hingga Rp72 miliar untuk aktivitas *influencer* guna meredam isu spesifik. Modus operandinya destruktif, mulai dari membanjiri kolom komentar dengan narasi seragam via akun *bot* demi menciptakan *trending topic*, hingga melakukan pembunuhan karakter (*character assassination*) terhadap kritisus, seperti kasus aktivis Ravio Patra dan komika Bintang Emon. Lebih jauh, strategi pengalihan isu (*distraction*) kerap diterapkan untuk menutupi blunder kebijakan, misalnya penggunaan tagar manipulatif dalam konflik agraria. Masyarakat dengan literasi rendah sering menjadi korban manipulasi *social proof* ini, tanpa sadar dieksplorasi untuk mengamplifikasi narasi hegemoni yang menyesatkan[3].

Urgensi deteksi *buzzer* semakin nyata pasca demonstrasi Agustus 2025. Studi kasus pada akun kreator Ferry Irwandi menunjukkan anomali signifikan. Saat menyuarakan fakta lapangan terkait korban penculikan dan penolakan eskalasi kerusuhan,

kolom komentarnya mengalami serangan masif dan terstruktur. Narasi pelabelan seperti "provokator" muncul secara seragam, berbeda jauh dari pola kritik organik. Serangan ini bertujuan mendekreditmasi fakta dan mengaburkan pelanggaran HAM, sekaligus mengindikasikan adanya orkestrasi rapi untuk membelokkan persepsi publik.

Berdasarkan eksplorasi awal, pendekatan *unsupervised learning* terbukti kurang efektif memisahkan *buzzer* politik di TikTok akibat kemampuan mimik yang tinggi. Oleh karena itu, penelitian ini beralih pada pendekatan *supervised learning* dengan mengomparasikan *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Secara empiris, metode tradisional ini mampu menyamai performa model *Deep Learning* (seperti CNN atau LSTM) pada klasifikasi teks pendek dengan efisiensi komputasi yang jauh lebih unggul (LR: 0,255 detik berbanding CNN: 119 detik)[4]. Studi *benchmarking* menempatkan SVM dan LR sebagai algoritma berkinerja tertinggi (*Rank 1 & 2*) untuk dataset tekstual. Fokus utama penelitian adalah mengisi kekosongan literatur melalui penerapan teknik *hybrid resampling* pada data tidak seimbang (*imbalanced*). Hal ini krusial karena bias data dapat menyebabkan pengabaian kelas minoritas, penerapan *oversampling* terbukti mampu mendongkrak *F1-Score* pelatihan secara signifikan (dari 0,84 menjadi 0,99), sehingga sistem diharapkan lebih sensitif mendekripsi komentar *buzzer*[5].

Tinjauan Pustaka

Penelitian mengenai deteksi akun *buzzer* dan analisis sentimen di media sosial telah berkembang pesat dengan menggunakan berbagai metode, baik itu *machine learning* maupun pendekatan berbasis kamus (*lexicon*). Meninjau penelitian-penelitian sebelumnya sangat penting untuk memahami kemajuan teknik yang digunakan serta mengetahui kekurangan dari metode yang sudah ada.

Lubis et al. (2025) melakukan deteksi akun *buzzer* pada platform X (Twitter) menggunakan algoritma *K-Nearest Neighbor* (KNN) yang dikombinasikan dengan teknik *resampling* SMOTE dan *Tomek Links* untuk menangani ketidakseimbangan data. Penelitian tersebut menghasilkan akurasi 91% pada rasio data 70:30 dengan nilai $K = 9$. Meskipun demikian, sebagai metode *instance-based learning*, KNN menghadapi tantangan komputasi yang tinggi pada data teks berdimensi besar. Selain itu, performa model sangat bergantung pada pemilihan nilai K , yang dinilai kurang efisien dibandingkan model berbasis *hyperplane* atau probabilistik untuk dataset yang kompleks[6].

Pada studi lain, Pramesti et al. (2025) mengidentifikasi *buzzer* pada ulasan produk menggunakan metode *Lexicon-Based* yang dievaluasi

dengan *Support Vector Machine* (SVM). Walaupun mencatat akurasi 93%, evaluasi mendalam menunjukkan kelemahan signifikan pada metrik *Recall* kelas *buzzer* yang hanya mencapai 0,50. Rendahnya nilai ini mengindikasikan kegagalan metode berbasis leksikon dalam menangkap makna implisit atau sarkasme, sehingga menyebabkan tingginya angka *false negative*[7].

Pendekatan *ensemble learning* diterapkan oleh Manullang et al. (2023) untuk memprediksi sentimen pemilu menggunakan algoritma *Random Forest* dengan pelabelan awal *Lexicon-Based*. Studi ini mencapai akurasi 88% dan *F1-Score* 0,79. Namun, ketergantungan pada pelabelan leksikon berpotensi menimbulkan bias pada data latih karena kata di luar kamus tidak terkласifikasi dengan tepat. Selain itu, *Random Forest* rentan mengalami *overfitting* pada fitur teks yang sangat variatif tanpa penyetelan parameter yang ketat[8].

Permasalahan *imbalanced data* pada Instagram ditangani oleh Wicaksono dan Cahyono (2024) melalui integrasi *Decision Tree* dan teknik SMOTE. Kombinasi ini meningkatkan akurasi hingga 85% pada skenario pembagian data 90:10. Kendati teknik *oversampling* berhasil meningkatkan performa, algoritma *Decision Tree* memiliki kelemahan terkait instabilitas struktur pohon; perubahan kecil pada data latih dapat mengubah model secara drastis, yang berdampak negatif pada generalisasi data baru[9].

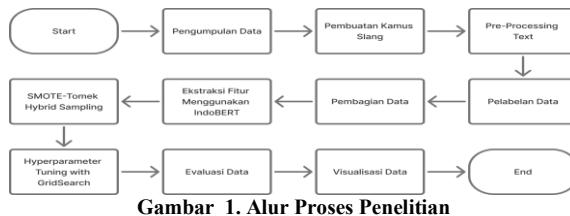
Dari perspektif atribut metadata, Perkasa et al. (2023) memfokuskan deteksi pada jumlah pengikut dan usia akun menggunakan *Gaussian Naive Bayes*, dengan capaian akurasi 80% dan AUC 0,78. Akan tetapi, pendekatan yang hanya berbasis metadata memiliki keterbatasan fatal karena mengabaikan konteks semantik narasi. Hal ini menjadi celah signifikan mengingat *buzzer* modern semakin canggih memanipulasi profil agar terlihat organik, sehingga analisis mendalam terhadap konten teks mutlak diperlukan[10].

Berdasarkan keterbatasan penelitian terdahulu mulai dari rendahnya *Recall* pada metode leksikon, instabilitas model pohon, hingga pengabaian konteks semantik penelitian ini mengusulkan pendekatan komparatif antara SVM dan *Logistic Regression* (LR) yang diperkuat dengan ekstraksi fitur *IndoBERT*. Metode ini bertujuan menangkap konteks kalimat secara mendalam pada komentar TikTok yang memiliki karakteristik linguistik unik. Penerapan SMOTE-Tomek juga dilakukan untuk menangani rasio *imbalance* yang ekstrem, dengan target menghasilkan model yang unggul pada metrik Akurasi, *Recall*, dan *F1-Score* untuk meminimalisasi kesalahan deteksi pada aspirasi publik.

Metode Penelitian

Kerangka Penelitian

Untuk mencapai tujuan penelitian, dilakukan serangkaian tahapan. Arsitektur dan alur proses penelitian yang diusulkan diilustrasikan pada Gambar 1.



Pengumpulan Data

Data sekunder penelitian ini diperoleh melalui *scraping* komentar video TikTok akun "Ferry Irwandi" terkait isu penculikan mahasiswa menggunakan layanan Apify. TikTok dipilih secara purposif karena signifikansinya sebagai media strategis komunikasi politik yang efektif membentuk persepsi publik melalui interaksi dua arah[11]. Data dikumpulkan dalam format CSV yang memuat ID pengguna dan teks komentar untuk mendeteksi aktivitas *buzzer*.

Pre-processing Data

Tahapan *preprocessing* krusial untuk memastikan kualitas data sebelum pemodelan, mengingat data media sosial mengandung *noise* (simbol, *typo*, *slang*) yang dapat menurunkan akurasi klasifikasi secara signifikan[12]. Penelitian ini menerapkan dua pendekatan utama: konstruksi kamus normalisasi berbasis AI dan *pipeline* pembersihan teks standar.

a. Pembuatan Kamus Slang

Kamus normalisasi disusun secara otomatis dengan mengidentifikasi kata-kata unik dari dataset gabungan yang tidak terdaftar dalam KBBI maupun *stopword*. Kandidat kata *slang* tersebut diproses menggunakan *Large Language Model* (LLM) Google Gemini melalui teknik *batch processing*.

Penggunaan LLM terbukti efektif menangani normalisasi teks via pendekatan *few-shot learning*, meminimalkan ketergantungan pada aturan manual yang kaku [13]. Model diinstruksikan menerjemahkan singkatan, *typo*, dan bahasa daerah ke bentuk baku Bahasa Indonesia, namun tetap mempertahankan entitas nama. Hasil pemetaan (misalnya: "bgt" → "banget") disimpan dalam format JSON untuk tahap normalisasi.

b. Text Cleaning

Tahap ini membersihkan elemen teks tak relevan menggunakan metode *regular expression* (RegEx), meliputi:

1. Penghapusan Entitas Digital: Menghapus URL, *mention*, dan *hashtag* yang tidak merepresentasikan sentimen emosional.
2. Penghapusan Karakter Non-ASCII: Menghilangkan emoji dan karakter asing agar teks hanya terdiri dari karakter standar.
3. Penghapusan Angka dan Tanda Baca: Mengganti angka dan tanda baca dengan spasi untuk mencegah penggabungan kata yang tidak diinginkan (misalnya: "yg.suka" menjadi "yg suka").

c. Case Folding

Case folding mengubah seluruh huruf menjadi *lowercase* demi menjaga konsistensi data. Langkah ini memastikan variasi kapitalisasi (seperti "Makan" dan "makan") dianggap sebagai token yang sama, sehingga efektif mengurangi dimensi fitur pada model[14].

d. Whitespace Removal

Tahap ini menghapus spasi berlebih (*extra spaces*) serta melakukan *trimming* di awal dan akhir kalimat. Tujuannya adalah merapikan struktur teks sebelum proses tokenisasi dilakukan.

e. Tokenizing dan Normalisasi Slang

Kalimat dipecah menjadi unit kata terpisah (*token*), langkah fundamental dalam pemrosesan teks[15]. Setiap token kemudian dicocokkan dengan kamus *slang* yang telah dibuat sebelumnya. Jika ditemukan (misalnya "gw"), token diganti dengan bentuk bakunya ("saya"). Proses ini krusial untuk mengatasi tingginya variasi penulisan di media sosial.

f. Filtering

Sebagai langkah akhir, dilakukan penyaringan untuk memastikan hanya token berkarakter alfabet murni (*alphabetic*) yang dipertahankan. Token yang mengandung angka atau simbol sisa dibuang untuk mereduksi dimensi fitur yang tidak perlu.

Pelabelan Data

Pelabelan data dilakukan secara manual menggunakan *tools* Label Studio untuk mengklasifikasikan komentar menjadi dua kategori: *Buzzer* dan *Non-Buzzer*. Penentuan label *Buzzer* didasarkan pada karakteristik akun yang menyebarkan pesan kampanye negatif (*negative campaign*) serta menggunakan identitas anonim (*sockpuppet*) dengan pola aktivitas *hit and run*[11].

Ekstraksi Fitur IndoBERT

Penelitian ini menggunakan model IndoBERT (Indonesian BERT) versi BERT-base-indonesian-522M sebagai metode ekstraksi fitur untuk mengubah data teksual menjadi representasi vektor. IndoBERT merupakan model bahasa pra-latih (Pre-trained Language Model) berbasis arsitektur Transformer yang dilatih menggunakan korpus Bahasa Indonesia berskala besar (Indo4B) untuk memahami nuansa morfologi dan sintaksis lokal[16].

Penanganan Ketidakseimbangan Data

Metode hybrid resampling SMOTE-Tomek dipilih untuk memitigasi bias kelas mayoritas yang umum terjadi pada pendekatan konvensional, sekaligus mengatasi kelemahan SMOTE standar dengan memperbaiki batasan keputusan (decision boundary) antar-kelas dan mencegah overfitting[17].

Secara mekanisme, algoritma bekerja dalam dua tahap sekvensial. Pertama, Synthetic Minority Over-sampling Technique (SMOTE) membangkitkan sampel sintetis kelas minoritas melalui interpolasi k-nearest neighbors menggunakan persamaan[18]:

$$x_{new} = x_i + (x_{zi} - x_i) \times \delta$$

Keterangan:

- x_{new} : Sampel sintetis baru.
- x_i : Sampel asli kelas minoritas (titik acuan).
- x_{zi} : Salah satu dari k -tetangga terdekat dari x_i .
- δ : Vektor acak bernilai 0 hingga 1.

Kedua, metode Tomek Links diterapkan untuk membersihkan data dengan menghapus pasangan tetangga terdekat (x_i, x_j) dari kelas berbeda. Langkah ini bertujuan menghilangkan noise serta mempertegas margin pemisah kelas yang tumpang tindih dalam ruang fitur vektor[19].

Implementasi dilakukan secara eksklusif pada data latih (training set) guna mencegah kebocoran data (data leakage) dan menjamin validitas evaluasi pada data uji. Strategi ini menargetkan keseimbangan distribusi kelas agar model mempelajari fitur linguistik selebih objek serta meminimalisir risiko false negative pada deteksi opini masyarakat.

Pemodelan Klasifikasi

Penelitian ini menerapkan pendekatan komparatif antara *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) untuk mengklasifikasikan fitur *embedding* IndoBERT berdimensi tinggi. Optimasi hiperparameter dilakukan menggunakan *GridSearch CV* untuk menentukan kombinasi parameter terbaik secara otomatis seperti tipe *kernel* pada SVM atau regularisasi pada LR berdasarkan skor F1 optimal guna menangani ketimpangan distribusi kelas[20].

Support Vector Machine (SVM) bekerja berdasarkan prinsip *Structural Risk Minimization* (SRM) untuk menemukan *hyperplane* pemisah terbaik yang memaksimalkan *margin* antar-kelas dalam ruang vektor[21]. Fungsi keputusan klasifikasi dirumuskan sebagai:

$$f(x) = \text{sign}(w \cdot x + b)$$

Keterangan:

- $f(x)$: Kelas prediksi (+1 untuk *Buzzer*, -1 untuk *Non-Buzzer*).
- w : Vektor bobot (*weight vector*) normal terhadap *hyperplane*.
- x : Vektor fitur input dari ekstraksi IndoBERT.
- b : Parameter bias pengatur posisi *hyperplane*.

Penelitian ini menguji efektivitas kernel *Linear* versus *Radial Basis Function* (RBF) untuk menangani kompleksitas semantik data[22]. Pemilihan SVM didasari oleh efektivitasnya dalam menangani *curse of dimensionality* pada keluaran model bahasa. SVM memiliki kemampuan generalisasi yang baik pada ruang fitur luas, sehingga diharapkan mampu membentuk *decision boundary* tegas untuk membedakan narasi organik dan *buzzer* tanpa mengalami *overfitting*[23].

Sebagai pembanding, digunakan *Logistic Regression* (LR) yang merupakan model *baseline* probabilistik. Berbeda dengan SVM, LR memprediksi peluang kelas menggunakan fungsi Sigmoid:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Keterangan:

- $\sigma(z)$: Probabilitas prediksi (rentang 0 hingga 1).
- z : Kombinasi linier input dan bobot ($w \cdot x + b$).
- e : Konstanta Euler (≈ 2.718).

LR dipilih karena efisiensi komputasi dan kemampuannya memberikan interpretasi probabilitas langsung (*probabilistic output*) yang krusial untuk analisis kekuatan prediksi[24].

Kedua algoritma dilatih menggunakan data yang telah diseimbangkan melalui teknik *hybrid resampling* SMOTE-Tomek untuk mencegah bias mayoritas. Evaluasi performa difokuskan pada metrik *Recall* dan *F1-Score* kelas minoritas guna memastikan sensitivitas model dalam mendeteksi opini organik, alih-alih sekadar mengejar akurasi global[25].

Evaluasi

Kinerja model dievaluasi menggunakan *Confusion Matrix* sebagai acuan dasar untuk memetakan detail kesalahan prediksi, menghindari bias yang sering muncul jika hanya mengandalkan akurasi pada data yang tidak seimbang. Mengingat karakteristik data yang *imbalanced*, fokus pengukuran ditekankan pada *Precision* untuk menjamin ketepatan prediksi agar tidak salah mendeteksi pengguna organik sebagai buzzer, serta *Recall* untuk mengetahui seberapa banyak buzzer asli yang berhasil dijaring oleh sistem. Keseimbangan antara kedua metrik tersebut diukur melalui *F1-Score* yang digunakan sebagai parameter utama penentu kualitas model. Selain evaluasi numerik, analisis performa diperlakukan menggunakan *Precision-Recall (PR) Curve* untuk melihat *trade-off* kinerja pada berbagai ambang batas (*threshold*), serta divalidasi secara kualitatif menggunakan *Word Cloud* untuk memastikan bahwa pola kata yang dideteksi model relevan dengan karakteristik narasi asli dari buzzer maupun non-buzzer.

Hasil dan Pembahasan

Penelitian ini menggunakan studi kasus pada kolom komentar akun TikTok Ferry Irwandi, khususnya pada unggahan video yang membahas isu penculikan mahasiswa pasca-demonstrasi Agustus 2025. Pemilihan objek ini didasarkan pada karakteristik konten kreator tersebut yang kerap memicu diskusi kritis, sehingga memunculkan interaksi intens antara opini organik masyarakat dan narasi yang dibangun oleh akun *buzzer*.

Proses akuisisi data dilakukan menggunakan bantuan Apify sebagai *web scraper*. Mengingat karakteristik data mentah media sosial yang memiliki banyak *noise* seperti *spam*, penggunaan *emoticon* berlebih, hingga komentar irelevan tahapan pembersihan data (*data cleaning*) menjadi krusial agar data dapat diproses secara optimal oleh model.

Setelah melalui proses penyaringan dan pelabelan manual (*manual labeling*), diperoleh dataset bersih sebanyak 1.502 komentar. Jumlah ini dinilai representatif untuk menggambarkan pola interaksi pada topik spesifik tersebut. Sebagaimana dirincikan pada Tabel 1, distribusi data memperlihatkan ketimpangan (*imbalance*) yang signifikan antar-kelas, yang merefleksikan tingginya intensitas aktivitas *buzzer* dalam isu ini.

Tabel 1. Persebaran Data

Label Sentimen	Jumlah Data	Persentase (%)
Buzzer	1.238	82,42%
Non-Buzzer	264	17,58%
Total	1.502	100%

Data tersebut memperlihatkan disparitas proporsi yang signifikan antara komentar yang terindikasi *Buzzer* dan komentar organik (*Non-Buzzer*). Jika ditinjau lebih dalam, kelas *Buzzer* mendominasi hampir 83% dari total data. Fenomena ini relevan dengan konteks sosial-politik, di mana aktivitas *buzzer* sering kali bertujuan melakukan *flooding* untuk membanjiri kolom komentar. Namun, dalam konteks *machine learning*, ketimpangan ini menimbulkan tantangan serius.

Tingkat ketidakseimbangan data diukur menggunakan *Imbalance Ratio* (IR) dengan membandingkan jumlah kelas minoritas terhadap mayoritas (N_{min}/N_{maj}). Perhitungan pada dataset ini menghasilkan nilai IR sebesar 0,2132, yang mengindikasikan kondisi *moderate to high imbalance*. Apabila data ini digunakan tanpa intervensi, model klasifikasi berpotensi mengalami bias terhadap kelas mayoritas (*majority class bias*). Model akan cenderung memprediksi setiap data baru sebagai *Buzzer* karena probabilitas statistik yang lebih besar, sehingga gagal menangkap pola unik pada kelas *Non-Buzzer*.

Dalam kasus deteksi *buzzer*, kesalahan ini berisiko tinggi karena dapat meningkatkan *false negative*, di mana aspirasi publik yang asli tertutup oleh prediksi mayoritas. Oleh karena itu, penanganan struktur data melalui teknik *resampling* menjadi langkah mutlak yang akan dibahas pada bagian selanjutnya.

Sebelum memasuki tahap pemodelan, analisis karakteristik data khususnya panjang komentar sangat diperlukan mengingat penggunaan IndoBERT yang memiliki batasan *input* atau *max sequence length* sebesar 512 token. Jika input melebihi batas ini, model akan melakukan pemotongan (*truncation*) yang berpotensi menghilangkan konteks penting.

Berdasarkan analisis statistik pada Tabel 2, karakteristik komentar TikTok berbeda signifikan dibandingkan *platform* lain seperti Facebook atau portal berita. Komentar cenderung singkat, padat, dan reaktif. Rata-rata panjang komentar dalam dataset ini hanya 17 token, dengan komentar terpanjang mencapai 151 token. Angka ini masih jauh di bawah batas maksimum IndoBERT, sehingga risiko hilangnya informasi akibat pemotongan teks dapat diminimalisir.

Tabel 2. Statistik Panjang Token

Metrik Statistik	Nilai (Token)
Rata-rata (Mean)	17,01
Standar Deviasi (Std)	16,03
Minimum	3
Median (50%)	12
Persentil 90%	33
Persentil 99%	86
Maksimum (Max)	151

Setelah verifikasi panjang data, sebanyak 1.502 komentar diproses menggunakan *tokenizer* bawaan IndoBERT. Keunggulan utama penelitian ini terletak pada penggunaan arsitektur BERT 12-layer yang mampu memahami konteks semantik secara mendalam (*contextual embedding*). Berbeda dengan metode statistik konvensional, IndoBERT dapat membedakan makna kata yang memiliki ambiguitas berdasarkan konteks kalimatnya. Sebagai ilustrasi, kata "bisa" dalam kalimat "ular memiliki bisa" dan "aku bisa makan" akan menghasilkan representasi vektor yang berbeda.

Proses ekstraksi fitur difokuskan pada token khusus [CLS] yang terdapat pada *last hidden state*. Token ini berfungsi sebagai representasi global (*sentence embedding*) dari keseluruhan input teks. Proses ini menghasilkan matriks fitur berukuran 1.502 x 768. Dengan demikian, setiap komentar kini terkonversi menjadi vektor padat (*dense vector*) dengan dimensi 768 yang kaya informasi dan siap untuk diproses oleh algoritma klasifikasi.

Untuk keperluan evaluasi, dataset dibagi dengan rasio 80:20, di mana 80% dialokasikan sebagai data latih (*training set*) dan 20% sebagai data uji (*testing set*). Rasio ini dipilih untuk memaksimalkan proses pembelajaran pada jumlah data yang terbatas, namun tetap mempertahankan jumlah sampel yang valid (301 data) untuk pengujian.

Mengingat kondisi data yang tidak seimbang (*imbalanced*), penggunaan teknik *Simple Random Sampling* berisiko menyebabkan distribusi kelas yang bias pada salah satu himpunan data. Untuk memitigasi risiko tersebut, penelitian ini menerapkan metode *Stratified Random Sampling*. Teknik ini memastikan proporsi kelas *Buzzer* dan *Non-Buzzer* pada data latih maupun data uji tetap konsisten dengan populasi aslinya, yakni pada kisaran 82% dan 17% bisa dilihat pada Tabel 3. Konsistensi distribusi ini menjamin bahwa akurasi model yang dihasilkan merefleksikan performa yang objektif pada kedua kelas.

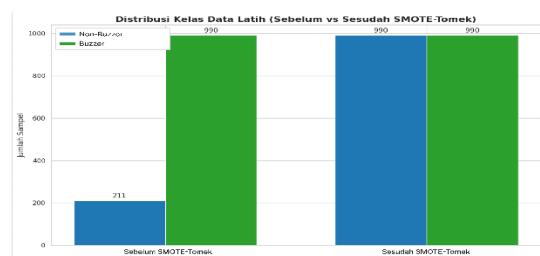
Tabel 3. Pembagian Data

Himpunan Data (Subset)	Jumlah Data	Non-Buzzer (0)	Buzzer (1)
Data Latih (Train)	1.201	211 (17,57%)	990 (82,43%)
Data Uji (Test)	301	53 (17,61%)	248 (82,39%)
Total	1.502	264 (17,58%)	1.238 (82,42%)

Mengingat nilai *Imbalance Ratio* (IR) sebesar 0,2132, penggunaan data latih tanpa intervensi berisiko menurunkan performa model pada kelas minoritas. Algoritma pembelajaran mesin secara alami cenderung meminimalkan *global error rate*, yang sering kali dicapai dengan mengorbankan akurasi

pada kelas minoritas demi memaksimalkan akurasi total. Hal ini dapat menghasilkan metrik evaluasi yang semu (*misleading accuracy*).

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan strategi *hybrid resampling* menggunakan metode SMOTE-Tomek. Prosedur ini secara spesifik hanya diterapkan pada data latih (*training set*), sedangkan data uji (*test set*) dibiarkan dalam kondisi aslinya. Pemisahan perlakuan ini merupakan langkah krusial untuk mencegah kebocoran data (*data leakage*) dan memastikan model dievaluasi menggunakan data riil, bukan data sintetis. Dampak perubahan distribusi data akibat proses ini dapat dilihat pada Gambar 2.



Gambar 2. Penyeimbangan Kelas

Melalui proses *resampling*, rasio distribusi kelas pada data latih berhasil diseimbangkan dari semula 1:4,7 menjadi 1:1, dengan masing-masing kelas memiliki 990 sampel. Keseimbangan distribusi ini mengeliminasi kecenderungan bias model terhadap kelas mayoritas. Dengan demikian, algoritma klasifikasi didorong untuk mempelajari fitur linguistik secara objektif guna membedakan karakteristik antara komentar organik dan komentar yang terindikasi sebagai *buzzer*.

Analisis Hasil Klasifikasi

Evaluasi performa *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) pada 301 data uji menunjukkan hasil yang sebanding, dengan akurasi global mencapai 0,89. Meskipun angka ini tampak memuaskan, penggunaan metrik akurasi pada *imbalanced dataset* dapat memberikan interpretasi yang bias (*accuracy paradox*). Komposisi data uji didominasi oleh 248 sampel kelas *Buzzer* (majoritas) dibandingkan 53 sampel kelas *Non-Buzzer* (minoritas). Sebagai ilustrasi, model yang memprediksi seluruh data sebagai kelas mayoritas (*majority class classifier*) secara otomatis akan menghasilkan akurasi dasar (*baseline accuracy*) sebesar 82%. Oleh karena itu, akurasi global tidak dapat dijadikan satu-satunya tolak ukur validitas model.

Evaluasi yang lebih komprehensif dilakukan menggunakan metrik *Precision*, *Recall*, dan *F1-Score* untuk mengukur sensitivitas model, khususnya dalam mendeteksi kelas minoritas. Hasil pengujian menunjukkan kedua model memiliki performa optimal pada kelas *Buzzer*. Hal ini didukung oleh

ketersediaan data latih yang memadai, memungkinkan SVM dan LR mengidentifikasi pola linguistik pada kelas tersebut dengan efektif. Sebagaimana ditunjukkan pada Tabel 4, nilai *Precision* dan *Recall* untuk kelas mayoritas konsisten berada di atas 0,92.

Tabel 4. Perbandingan Hasil Klasifikasi

Model	Kelas /Metrik	Precision	Recall	F1-Score
SVM	Non-Buzzer (0)	0,68	0,72	0,70
	Buzzer (1)	0,94	0,93	0,93
	Macro	0,81	0,82	0,82
	Average Accuracy			0,89
Logistic Regression	Non-Buzzer (0)	0,67	0,79	0,72
	Buzzer (1)	0,95	0,92	0,93
	Macro	0,81	0,85	0,83
	Average Accuracy			0,89

Ujian krusial bagi model klasifikasi terletak pada sensitivitasnya dalam mengidentifikasi kelas minoritas. Saat dievaluasi pada kelas komentar organik (*Non-Buzzer*), performa kedua model mengalami penurunan dengan karakteristik yang berbeda. SVM menunjukkan tendensi yang lebih konservatif, tercermin dari nilai *Recall* sebesar 0,72 dan *Precision* 0,68. Pendekatan ini mengindikasikan bahwa meskipun SVM cukup selektif dalam prediksi positif, model ini memiliki *False Negative Rate* yang cukup tinggi, di mana sekitar 28% data *Non-Buzzer* gagal terdeteksi.

Sebaliknya, *Logistic Regression* (LR) menunjukkan karakteristik yang lebih inklusif. Meskipun nilai *Precision*-nya (0,67) sedikit di bawah SVM, LR berhasil meningkatkan *Recall* secara signifikan mencapai 0,79. Dalam konteks analisis opini publik, tingginya nilai *Recall* menjadi prioritas utama untuk meminimalisir risiko hilangnya aspirasi murni masyarakat akibat kesalahan klasifikasi (*misclassification*).

Berdasarkan analisis komparatif tersebut, penelitian ini menetapkan *Logistic Regression* sebagai model terbaik. Keputusan ini didukung oleh superioritas skor *F1-Score* pada kelas minoritas (0,72 vs 0,70) serta *Macro Average Recall* (0,85 vs 0,82).

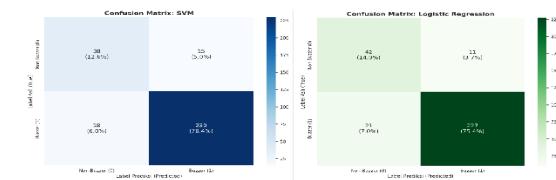
Secara teoritis, keunggulan LR dapat diatribusikan pada penggunaan fungsi *sigmoid* yang bersifat probabilistik. Hal ini memungkinkan pembentukan *decision boundary* yang lebih fleksibel dan adaptif terhadap distribusi data. Sementara itu, performa SVM terkendala oleh prinsip *margin maximization*. Pada dataset dengan ketimpangan tinggi

(*imbalanced*), *hyperplane* yang dibentuk SVM cenderung terdesak oleh kelas mayoritas, sehingga mempersempit area deteksi bagi kelas minoritas dan menurunkan sensitivitas model.

Evaluasi Confusion Matrix

Evaluasi kinerja model dalam penelitian ini tidak dapat hanya bertumpu pada metrik akurasi global (*accuracy*). Mengingat karakteristik kasus deteksi *buzzer* yang sensitif, analisis mendalam menggunakan *Confusion Matrix* diperlukan untuk memahami pola distribusi kesalahan prediksi (*prediction error*) pada algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Analisis ini bertujuan untuk memetakan karakteristik model, khususnya terkait kecenderungan terjadinya *False Positive* atau *False Negative*.

Dalam konteks klasifikasi ini, terminologi evaluasi didefinisikan sebagai berikut: *True Negative* (TN) merepresentasikan komentar organik yang berhasil diidentifikasi dengan benar. *False Positive* (FP) merupakan kesalahan klasifikasi di mana komentar organik secara keliru diprediksi sebagai *buzzer* sebuah kesalahan yang berisiko merugikan pengguna asli. Sebaliknya, *False Negative* (FN) adalah kondisi kegagalan sistem dalam mendeteksi keberadaan *buzzer*, sedangkan *True Positive* (TP) menunjukkan keberhasilan model dalam mengidentifikasi target *buzzer* secara tepat. Visualisasi hasil *Confusion Matrix* dari kedua model disajikan pada Gambar 3.



Gambar 3. Confusion Matrix

Analisis terhadap hasil *Support Vector Machine* (SVM) memperlihatkan karakteristik pembentukan *decision boundary* yang tegas. SVM mencatatkan nilai *True Positive* (TP) sebesar 230, sedikit lebih tinggi dibandingkan LR. Secara teknis, angka ini mengindikasikan bahwa SVM memiliki sensitivitas yang tinggi dalam mengidentifikasi kelas target (*Buzzer*).

Namun, pendekatan *margin maximization* yang diterapkan SVM memiliki implikasi pada peningkatan kesalahan Tipe I (*Type I Error*). Upaya algoritma untuk memperlebar margin pemisah cenderung menyebabkan *overfitting* pada area perbatasan kelas, sehingga angka *False Positive* (FP) meningkat hingga 15 kasus. Dampak praktis dari tingginya FP ini adalah risiko kesalahan klasifikasi terhadap pengguna organik, yang berpotensi menyebabkan penghapusan komentar atau penangguhan akun yang tidak valid. Hal ini mengindikasikan bahwa *hyperplane* SVM terlalu ekspansif dalam memetakan wilayah kelas minoritas dan mengklasifikasikan *noise* sebagai fitur *buzzer*.

Sebaliknya, *Logistic Regression* (LR) menunjukkan pendekatan yang lebih seimbang. Meskipun nilai *True Positive* (227) terpaut 3 poin di bawah SVM, LR menunjukkan superioritas dalam meminimalisir kesalahan pada data organik. Model ini berhasil mengidentifikasi 42 *True Negative* (TN), lebih tinggi dibandingkan SVM yang hanya mencatat 38 TN.

Keunggulan ini berkorelasi dengan mekanisme kerja LR yang berbasis probabilitas. Dengan *threshold* standar 0,5, LR membentuk batas keputusan yang lebih fleksibel (*soft decision boundary*) dibandingkan SVM. Algoritma ini mengestimasi peluang keanggotaan kelas alih-alih sekadar memaksimalkan jarak geometris, sehingga menghasilkan prediksi yang lebih presisi bagi data pengguna asli. Hal ini dibuktikan dengan rendahnya angka *False Positive* (11 kasus) dan nilai *Specificity* yang lebih unggul, yakni 79,2% berbanding 71,7% milik SVM.

Dalam kondisi dataset yang sangat tidak seimbang (*imbalanced*) di mana proporsi kelas *buzzer* (248) mendominasi kelas *non-buzzer* (53) kemampuan model untuk mempertahankan nilai *Specificity* menjadi indikator validitas yang krusial. Rincian perbandingan performa kedua model berdasarkan parameter *Confusion Matrix* disajikan pada Tabel 5.

Tabel 5. Analisis Confusion Matrix

Fitur Evaluasi	Support Vector Machine (SVM)	Logistic Regression (LR)	Analisis Komparatif
Akurasi Total	89,04% (268/301)	89,37% (269/301)	LR sedikit lebih unggul secara global.
True Positive (TP)	230 (Sangat Tinggi)	227 (Tinggi)	SVM sedikit lebih agresif mendeteksi Buzzer.
True Negative (TN)	38 (Cukup)	42 (Sangat Baik)	LR lebih andal mengenali komentar asli.
False Positive (FP)	15 (Sedang)	11 (Rendah)	LR meminimalisir risiko "salah tuduh" pada user asli.
False Negative (FN)	18 (Rendah)	21 (Sedang)	SVM lebih ketat dalam menjaring Buzzer (minim kebocoran).
Sensitivitas (Recall)	92.7%	91.5%	SVM lebih unggul dalam menemukan seluruh Buzzer.
Spesifisitas	71.7%	79.2%	LR jauh lebih baik menangani kelas minoritas (Non-Buzzer).

Berdasarkan data pada Tabel 5, *Logistic Regression* (LR) menunjukkan superioritas marginal dengan

akurasi global sebesar 89,37%, berbanding 89,04% pada *Support Vector Machine* (SVM). Kendati selisih numerik tersebut tampak tidak signifikan, stabilitas LR dalam menangani ketimpangan data (*imbalanced dataset*) menegaskan karakteristiknya yang lebih robust.

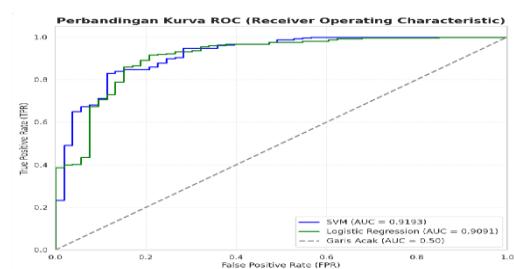
Defisiensi performa SVM pada kelas minoritas berkorelasi erat dengan prinsip *margin maximization*, yang cenderung bias terhadap kelas mayoritas. Sebaliknya, pendekatan probabilistik yang diterapkan LR terbukti lebih konsisten dalam menjaga generalisasi model. Oleh karena itu, dengan memprioritaskan mitigasi kesalahan deteksi pada pengguna organik (*false positive*), penelitian ini merekomendasikan *Logistic Regression* sebagai model yang lebih valid untuk implementasi jangka panjang.

Evaluasi Hasil Model

Evaluasi model yang hanya bertumpu pada metrik kuantitatif seperti akurasi atau *F1-score* belum memadai untuk memahami karakteristik pembelajaran model secara komprehensif. Meskipun metrik tersebut memberikan indikator performa, angka-angka tersebut tidak dapat memvisualisasikan mekanisme model dalam mendiskriminasi pola linguistik antara *buzzer* dan pengguna organik. Oleh karena itu, penelitian ini melengkapi analisis numerik dengan pendekatan visual untuk menginterpretasi perilaku model. Pendekatan ini meliputi penggunaan *Word Cloud* untuk memetakan distribusi narasi, kurva ROC untuk menguji reliabilitas prediksi, serta *t-Distributed Stochastic Neighbor Embedding* (t-SNE) untuk memvisualisasikan representasi vektor IndoBERT dalam ruang dimensi rendah.

Untuk mengukur keandalan model dalam memisahkan kelas *buzzer* dan *non-buzzer*, digunakan analisis *Receiver Operating Characteristic* (ROC) Curve. Kurva ini merepresentasikan hubungan komparatif antara *True Positive Rate* (sensitivitas) dan *False Positive Rate* (1-spesifitas).

Secara teoritis, semakin kurva mendekati sudut kiri atas (*top-left corner*), semakin tinggi kemampuan diskriminatif model dalam mengklasifikasikan data dengan tepat serta meminimalisir tingkat kesalahan (*false positive*). Perbandingan performa kurva ROC antara model SVM dan *Logistic Regression* dapat dilihat pada Gambar 4.



Gambar 4. Kurva ROC

Baik *Support Vector Machine* (SVM) maupun *Logistic Regression* (LR) yang diintegrasikan dengan fitur IndoBERT menunjukkan performa yang solid, dengan nilai *Area Under Curve* (AUC) melampaui 0,90. Capaian ini tergolong signifikan mengingat karakteristik data teks media sosial yang umumnya tidak terstruktur, memiliki *noise* tinggi, dan banyak menggunakan *slang*.

Berdasarkan analisis komparatif, SVM menunjukkan stabilitas yang sedikit lebih unggul dengan skor AUC 0,9193, melampaui LR yang mencatatkan skor 0,9091.

Selisih performa yang marginal ini mengindikasikan bahwa faktor determinan utama terletak pada kualitas fitur IndoBERT. Representasi vektor (*embedding*) yang dihasilkan IndoBERT terbukti sangat *robust*, sehingga mampu menunjang kinerja berbagai algoritma klasifikasi secara optimal. Meskipun demikian, kurva SVM yang lebih dominan merefleksikan kemampuan algoritma ini dalam membentuk *margin* yang lebih aman saat menangani data dengan probabilitas ambigu. Berdasarkan metrik AUC, kombinasi IndoBERT dan SVM merupakan konfigurasi yang paling optimal secara statistik. Selanjutnya, distribusi kata dominan pada setiap kelas divisualisasikan melalui *Word Cloud* pada Gambar 5.



Gambar 5. Wordcloud

Visualisasi *Word Cloud* merepresentasikan distribusi terminologi dominan yang menjadi karakteristik utama narasi kedua kelas. Pada kelompok *Buzzer*, diskursus yang terbangun cenderung bersifat politis, defensif, dan menyerang. Tingginya frekuensi kemunculan kata kunci seperti "salah", "blunder", "dihilangkan", "minta maaf", dan "klarifikasi" mengindikasikan pola narasi yang berorientasi pada delegitimasi lawan atau pengalihan isu (*counter-narrative*). Selain itu, penggunaan kata sapaan informal seperti "lu" dan "bang" secara masif mencerminkan gaya bahasa yang konfrontatif dan agresif.

Sebaliknya, karakteristik *Word Cloud* pada kelas *Non-Buzzer* memperlihatkan nuansa interaksi yang jauh lebih organik dan suportif. Terminologi yang mendominasi meliputi leksikon positif seperti "semangat", "keluarga", "alhamdulillah", "sehat", dan "semoga". Hal ini merefleksikan pola komunikasi wajar yang berbasis pada koneksi sosial dan dukungan emosional (*emotional support*), bukan penggiringan opini politik. Divergensi leksikal yang signifikan antara gaya bahasa konfrontatif *buzzer* dan gaya afektif pengguna organik inilah yang menjadi fitur distingtif utama (*discriminative feature*), yang

memungkinkan model membedakan kedua kelas dengan akurasi tinggi.

Kesimpulan dan Saran

Kesimpulan

Penelitian ini menyimpulkan bahwa integrasi fitur ekstraksi IndoBERT dengan teknik *hybrid resampling* SMOTE-Tomek terbukti efektif dalam mengatasi ketidakseimbangan data ekstrem pada deteksi *buzzer* di TikTok, menghasilkan model yang *robust* dengan *Area Under Curve* (AUC) melampaui 0,90. Meskipun *Support Vector Machine* (SVM) unggul secara marginal dalam sensitivitas deteksi *buzzer* (AUC 0,9193), *Logistic Regression* (LR) direkomendasikan sebagai model terbaik karena menunjukkan superioritas signifikan dalam metrik *specificity* (79,2%) dan F1-Score kelas minoritas (0,72). Karakteristik probabilistik LR terbukti lebih adaptif dalam meminimalisir kesalahan Tipe I (*False Positive*), menjadikannya pendekatan yang lebih valid dan etis untuk menjaga orisinalitas aspirasi publik dibandingkan SVM yang cenderung agresif dalam memperluas margin keputusan hingga mengorbankan akurasi pada data organik.

Saran

Berdasarkan temuan dan keterbatasan riset, penelitian selanjutnya disarankan untuk memperluas cakupan dataset dengan melibatkan variasi topik politik yang lebih beragam dan lintas-kreator guna menguji generalisasi model terhadap perubahan narasi yang dinamis. Selain itu, pengembangan sistem deteksi sebaiknya tidak hanya bertumpu pada fitur semantik tekstual semata, melainkan perlu mengintegrasikan pendekatan multimodal yang menggabungkan analisis jejaring sosial (*Social Network Analysis*), metadata profil, dan pola perilaku temporal (*time-series behavior*) untuk mendeteksi *buzzer* yang semakin adaptif. Penerapan teknik *Explainable AI* (XAI) juga sangat direkomendasikan di masa depan untuk memberikan transparansi interpretasi keputusan model, sehingga pelabelan akun dapat dipertanggungjawabkan secara logis dan mengurangi potensi bias algoritma dalam ekosistem digital.

Daftar Pustaka

- [1] V. Tanuwijaya and S. C. M, "FENOMENA JINGLE " OKE GAS ": STUDI NETNOGRAFI KAMPANYE POLITIK DI TIKTOK Abstrak " OKE GAS " JINGLE PHENOMENON: A NETNOGRAPHIC STUDY OF," vol. 8, no. 2, pp. 165–184, 2024.
 - [2] M. R. Hasanah, F. A. Putri, P. K. Nisa, U. Islam, N. Syarif, and H. Jakarta, "Retorika Politik Anies Baswedan pada Kampanye Pemilu 2024 Terhadap Persepsi Publik di TikTok," vol. 5, no. 2, pp. 566–572, 2025.
 - [3] P. J. Widodo, "Penggunaan Buzzer dalam Hegemoni Pemerintahan Joko Widodo," vol.

- [4] 1, no. 2, pp. 70–81, 2022.
- [5] K. S. Id, M. Girnyk, and L. D. Id, “Short text classification with machine learning in the social sciences : The case of climate change on Twitter,” pp. 1–26, 2023, doi: 10.1371/journal.pone.0290762.
- [6] S. F. Taskiran, B. Turkoglu, E. Kaya, and T. Asuroglu, “OPEN A comprehensive evaluation of oversampling techniques for enhancing text classification performance,” pp. 1–20, 2025.
- [7] A. A. Lubis, S. Iskandar, A. Idrus, Z. Indra, and K. S. S, “Klasifikasi Akun Buzzer Menggunakan Algoritma K-Nearest Neighbor pada Tagar # STYTanpaDiasporaNol di Media Sosial X,” 2025.
- [8] A. D. Pramesti, K. Umam, and M. R. Handayani, “Identification of Buzzers in Skincare Reviews Using a Lexicon-Based Sentiment Analysis Method,” vol. 9, no. 5, pp. 2598–2606, 2025.
- [9] O. Manullang, C. Prianto, and N. H. Harani, “Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest,” *J. Ilm. Inform.*, vol. 11, no. 02, pp. 159–169, 2023, doi: 10.33884/jif.v11i02.7987.
- [10] N. Cahyono, S. Informatika, and U. A. Yogyakarta, “Analisis Sentimen Komentar Instagram Pada Program Kampus,” vol. 8, no. 2, pp. 2372–2381, 2024.
- [11] Catur Arpal Perkasa, Amalia Andjani Arifiyanti, and Agus Salim, “Klasifikasi Akun Buzzer Pada Twitter Menggunakan Algoritma Naive Bayes,” *J. Ilm. Tek. Inform. dan Komun.*, vol. 3, no. 1, pp. 01–12, 2023, doi: 10.55606/juitik.v3i1.363.
- [12] R. Mustika, “Pergeseran Peran Buzzer Ke Dunia Politik Di Media Sosial,” *Diakom J. Media dan Komun.*, vol. 2, no. 2, pp. 144–151, 2019, doi: 10.17933/diakom.v2i2.60.
- [13] N. Nyoman, E. Perimawati, R. R. Huizen, and D. P. Hostiadi, “Analisa Pengaruh Pre-Procesing Data Untuk Model Deteksi Akun Palsu Pada Media Sosial,” vol. 2, no. 1, pp. 1117–1122, 2025.
- [14] M. Wong, A. Alshehri, S. Kao, and H. He, “PolyNorm : Few-Shot LLM-Based Text Normalization for Text-to-Speech,” pp. 77–85, 2025.
- [15] U. Khairani *et al.*, “PENGARUH TAHAPAN PREPROCESSING TERHADAP MODEL INDOBERT DAN THE INFLUENCE OF PREPROCESSING STAGES ON INDOBERT AND INDOBERTWEET MODELS FOR EMOTION DETECTION IN INSTAGRAM NEWS,” vol. 11, no. 4, 2024, doi: 10.25126/jtiik.1148315.
- [16] B. Wilie *et al.*, “IndoNLU : Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” pp. 843–857, 2020.
- [17] M. C. Monard, “A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data”.
- [18] R. Peranganin, E. J. G. Harijanja, I. K. Jaya, and B. Rumahorbo, “Penerapan Algoritma Safe-Level-Smote Untuk Peningkatan Nilai G-Mean Dalam Klasifikasi Data Tidak Seimbang,” *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 4, no. 1, pp. 67–72, 2020, doi: 10.46880/jmika.vol4no1.pp67-72.
- [19] E. F. Swana and W. Doorsamy, “Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset,” 2022.
- [20] C. Hsu, C. Chang, and C. Lin, “A Practical Guide to Support Vector Classification,” vol. 1, no. 1, pp. 1–20, 2025.
- [21] L. Saitta, “Support-Vector Networks,” vol. 297, pp. 273–297, 1995.
- [22] A. Roy and S. Chakraborty, “Support vector machine in structural reliability analysis: A review,” *Reliab. Eng. Syst. Saf.*, vol. 233, no. April, 2023, doi: 10.1016/j.ress.2023.109126.
- [23] F. A. Wicaksono, A. Romadhyony, and Hasmawati, “Sentiment Analysis of University Social Media Using Support Vector Machine and Logistic Regression Methods,” *Ind. J. Comput.*, vol. 7, no. 2, pp. 15–24, 2022, doi: 10.34818/indojc.2022.7.2.638.
- [24] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, “Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling,” *J. Inf. Syst. Informatics*, vol. 5, no. 3, pp. 915–927, 2023, doi: 10.51519/journalisi.v5i3.539.
- [25] E. C. Zabor, C. A. Reddy, R. D. Tendulkar, and S. Patil, “Logistic Regression in Clinical Studies,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 112, no. 2, pp. 271–277, 2022, doi: 10.1016/j.ijrobp.2021.08.007.