

Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang

Adinda Anas Qolbu¹, Nina Fitriyati^{2*}, Nur Inayah³

^{1,2,3}Program Studi Matematika, Fakultas Sains dan Matematika, Universitas Islam Negeri Syarif Hidayatullah, Jl. Ir. H. Djuanda No. 95 Ciputat, Kota Tangerang Selatan 15412, Indonesia

Korespondensi; Nina Fitriyati, Email: nina.fitriyati@uinjkt.ac.id

Abstrak

Penelitian ini bertujuan untuk membandingkan performa tiga pendekatan analisis sentimen, yaitu Naïve Bayes, Support Vector Machine (SVM), dan Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT), pada layanan IndiHome menggunakan data Twitter. Keterbatasan model tradisional melatarbelakangi penelitian ini dalam mengenali opini positif dan tantangan ketidakseimbangan data yang sering muncul dalam analisis berbasis media sosial. Data penelitian berupa 7393 tweet (Januari 2019–Agustus 2024) yang dilabeli secara manual menjadi sentimen positif dan negatif. Model dievaluasi menggunakan *stratified 10-fold cross validation* dan data uji, dengan penerapan teknik penanganan ketidakseimbangan berupa Synthetic Minority Oversampling Technique (SMOTE) dan pembobotan kelas (*class weighting*). Hasil menunjukkan IndoBERT unggul dengan akurasi 0,96 dan F1-score makro 0,95 tanpa penanganan khusus, sedangkan SVM mencapai akurasi 0,95 dengan pembobotan kelas, dan Naïve Bayes meningkat dari akurasi 0,89 menjadi 0,92 setelah SMOTE. Analisis tren sentimen menunjukkan opini negatif mendominasi, terutama terkait kecepatan dan kestabilan layanan. Temuan ini menegaskan bahwa IndoBERT lebih efektif dalam memahami konteks bahasa Indonesia, sementara teknik penanganan data tetap relevan untuk meningkatkan performa model tradisional. Hasil penelitian ini penting karena memberikan dasar empiris dalam pemilihan model analisis sentimen yang lebih akurat, adaptif terhadap bahasa Indonesia, dan bermanfaat dalam meningkatkan kualitas layanan.

Kata Kunci: Analisis Sentimen, *Stratified 10-Fold Cross Validation*, SMOTE, Pembobotan Kelas.

Abstract

This study aims to compare the performance of three sentiment analysis approaches, namely Naïve Bayes, Support Vector Machine (SVM), and Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT), on IndiHome services using Twitter data. The limitations of traditional models underlie this study in recognizing positive opinions and the challenge of data imbalance that often arises in social media based analysis. The research data consist of 7,393 tweets (January 2019–August 2024) manually labeled into positive and negative sentiments. Models were evaluated using stratified 10-fold cross validation and test data, with the application of imbalance handling techniques such as Synthetic Minority Oversampling Technique (SMOTE) and class weighting. Results show IndoBERT excels with 0.96 accuracy and 0.95 macro F1-score without special handling, while SVM reaches 0.95 accuracy with class weighting, and Naïve Bayes improves from 0.89 to 0.92 accuracy after SMOTE. Sentiment trend analysis indicates negative opinions dominate, mainly regarding speed and service stability. These findings confirm IndoBERT is more effective in understanding Indonesian context, while data handling remains relevant for improving traditional models. This study's results are important because they offer an empirical foundation for choosing sentiment analysis models that are more accurate, adaptive to Indonesian language, and useful for improving service quality.

Keywords: Sentiment Analysis, Stratified 10-Fold Cross Validation, SMOTE, Class Weighting.

Pendahuluan

Penggunaan internet di Indonesia terus meningkat dalam satu dekade terakhir, dengan jumlah pengguna mencapai 185 juta pada awal 2024 [1]. Seiring dengan meningkatnya kebutuhan internet, kualitas layanan penyedia jaringan menjadi faktor penting yang menentukan kepuasan pengguna. IndiHome, sebagai penyedia layanan internet terbesar dengan pangsa pasar 67,54% pada 2022 [2], menghadapi banyak keluhan terkait kecepatan, kestabilan koneksi, dan biaya layanan [3]. Opini publik tersebut tereksresi secara masif di media sosial, khususnya Twitter, sehingga analisis sentimen dapat menjadi sarana strategis untuk mengevaluasi persepsi pelanggan dan mendukung perbaikan layanan.

Dalam penelitian ini, Twitter dipilih karena merupakan salah satu media sosial yang berfungsi sebagai wadah untuk menampung dan mengekspresikan berbagai opini publik [4]. Selain itu, Format unggahannya yang berbasis teks memungkinkan proses pengumpulan dan analisis opini publik secara lebih terstruktur. Beberapa penelitian di Indonesia juga telah mengkaji penerapan analisis sentimen pada berbagai layanan publik dan perusahaan nasional dengan data yang bersumber dari media sosial Twitter, seperti studi terhadap layanan Gojek [5], BMKG Nasional [6], dan Tokopedia [7]. Penelitian-penelitian ini menunjukkan penerapan analisis sentimen yang luas pada berbagai sektor layanan digital di Indonesia. Hal ini sejalan dengan fokus penelitian ini yang menelaah persepsi pengguna terhadap layanan IndiHome.

Analisis sentimen berbasis *machine learning* telah banyak dilakukan dengan algoritma tradisional seperti Naïve Bayes dan SVM [8][9]. Namun, penelitian sebelumnya terhadap layanan IndiHome menunjukkan kelemahan pada nilai *recall* kelas positif yang rendah, masing-masing hanya 57% untuk Naïve Bayes dan 64% untuk SVM [10]. Hal ini mengindikasikan keterbatasan model tradisional dalam menangkap opini positif. Di sisi lain, munculnya model berbasis *transformer* seperti IndoBERT membuka peluang baru, karena mampu memahami konteks bahasa Indonesia secara lebih mendalam dan telah terbukti unggul dalam berbagai tugas *Natural Language Processing* [11][12].

Selain pemilihan model, tantangan lain yang sering muncul adalah ketidakseimbangan kelas data, di mana opini negatif cenderung lebih dominan dibanding positif [13]. Kondisi ini dapat menyebabkan bias model terhadap kelas mayoritas. Oleh karena itu, diperlukan penerapan strategi penanganan data tidak seimbang, seperti *Synthetic Minority Oversampling Technique* (SMOTE) atau pembobotan kelas (*class weighting*), agar model mampu mengenali distribusi kelas secara lebih adil.

Dengan mempertimbangkan keterbatasan model tradisional dan tantangan distribusi data yang tidak seimbang tersebut, diperlukan penelitian yang secara komprehensif membandingkan berbagai pendekatan untuk menemukan metode analisis sentimen yang paling efektif. Oleh karena itu, penelitian ini bertujuan untuk membandingkan performa Naïve Bayes, SVM, dan IndoBERT dalam analisis sentimen terhadap layanan IndiHome; serta menganalisis pengaruh penanganan ketidakseimbangan data terhadap kinerja masing-masing model. Dengan melibatkan pendekatan dari model sederhana hingga model berbasis *transformer*, penelitian ini diharapkan memberikan gambaran komprehensif terhadap efektivitas berbagai metode analisis sentimen pada konteks layanan internet Indonesia, sekaligus menawarkan rekomendasi pemilihan model yang lebih tepat bagi studi maupun praktik industri.

Landasan Teori

Multinomial Naïve Bayes merupakan algoritma klasifikasi berbasis probabilistik yang sering digunakan untuk data teks. Model ini mengasumsikan bahwa fitur saling independen satu sama lain, dan menghitung probabilitas setiap kelas berdasarkan frekuensi fitur dalam dokumen. Prinsip utamanya berdasarkan Teorema Bayes, yang dinyatakan sebagai berikut:

$$P(C_k|x) = \frac{P(x|C_k) \cdot P(C_k)}{P(x)}, \quad (1)$$

dengan $P(C_k|x)$ adalah probabilitas posterior dari suatu kelas C_k diberikan vektor fitur x , $P(x|C_k)$ adalah likelihood, yaitu probabilitas kemunculan fitur x jika diketahui kelasnya C_k atau dikenal sebagai probabilitas kondisional, $P(C_k)$ adalah probabilitas awal (*prior*) dari kelas C_k , dan $P(x)$ adalah likelihood marginal dari fitur x [14].

SVM adalah metode klasifikasi yang bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas data dengan margin maksimum. Model SVM bekerja dengan membangun fungsi keputusan $f(x)$, yang dirumuskan sebagai berikut:

$$f(x) = \text{sign}(\sum_{i=1}^N a_i y_i K(x_i, x) + b), \tag{2}$$

dengan $f(x)$ adalah fungsi keputusan untuk menentukan kelas dari input x , x_i adalah data latih, y_i adalah label kelas (positif atau negatif), a_i adalah koefisien hasil pelatihan, $K(x_i, x)$ Adalah fungsi kernel yang mengukur kemiripan antara x dan x_i , dan b adalah bias atau konstanta dalam fungsi keputusan [15].

Transformers adalah arsitektur neural network yang kuat untuk tugas NLP seperti text classification, language translation, dan language modelling. Berbeda dari neural network tradisional, transformers memproses data sebagai urutan token dan dapat menangani sequence panjang serta memahami hubungan antar-token yang berjauhan [16]. Mekanisme *attention* yang merupakan inti dari transformer, membuat model bisa memberi perhatian pada bagian-bagian tertentu dari urutan input saat menghasilkan output. Dengan kata lain, model tidak melihat semua kata secara sama rata, namun menyesuaikan fokusnya pada kata-kata yang lebih penting sesuai kebutuhan. Jika diketahui *query* (q), *key* (k), dan *value* (v), maka *attention* dihitung menggunakan rumus:

$$\text{Attention}(q, k, v) = \text{softmax}\left(\frac{q \cdot k^T}{\sqrt{d_k}}\right) v, \tag{3}$$

dengan d_k adalah ukuran (dimensi) dari *key* [17].

IndoBERT adalah model transformer berbasis BERT yang dilatih menggunakan dataset Indo4B. IndoBERT adalah model kontekstual Bahasa Indonesia yang cocok untuk tugas Natural Language Processing seperti analisis sentimen [18]. BERT (Bidirectional Encoder Representations from Transformers) di-pre-train pada korpus besar (BooksCorpus & Wikipedia) menggunakan dua unsupervised tasks: *Masked Language Modeling* (MLM), di mana 15 % token acak di-mask dan diprediksi ulang, serta *Next Sentence Prediction* (NSP), yang melatih model memahami hubungan antar-kalimat dengan membedakan kalimat berurutan atau tidak [19]. Setelah pre-training, model kemudian di-fine-tune (melanjutkan training) secara end-to-end dengan menambahkan satu output layer untuk berbagai *downstream tasks* seperti sentiment classification tanpa mengubah arsitektur inti, sehingga mampu mencapai performa unggul pada berbagai evaluasi NLP. Penelitian dengan IndoBERT ini diawali dengan memanfaatkan model pre-train indobenchmark/indobert-base-p1 sebagai basis untuk *fine-tuning*.

Bahan dan Metode

Data yang digunakan dalam penelitian ini berupa tweet yang dikumpulkan dari periode Januari 2019 hingga Oktober 2024 dengan kata kunci "indihome" untuk memastikan relevansi topik. Proses pengumpulan data memperhatikan aspek etika penelitian, termasuk menjaga kerahasiaan identitas pengguna dan mematuhi kebijakan privasi Twitter. Data mentah kemudian disimpan untuk diproses lebih lanjut, dengan total sebanyak 7393 data. Berikut adalah langkah-langkah yang digunakan untuk pengolahan data.

1. Preprocessing Data

Tahap ini dilakukan untuk mempersiapkan teks agar dapat diolah secara optimal oleh model klasifikasi. Langkah pertama adalah pembersihan teks, yaitu menghapus elemen-elemen yang tidak relevan seperti

URL, hashtag, mention, angka, simbol, dan emotikon. Selanjutnya, dilakukan *case folding* untuk mengonversi seluruh teks menjadi huruf kecil. Proses berikutnya adalah *slang normalization*, yaitu mengganti kata tidak baku atau singkatan ke bentuk baku. Setelah itu, *stopword removal* dilakukan untuk menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis. Tahap terakhir adalah *stemming* dan *lemmatization*, yaitu mengembalikan kata ke bentuk dasarnya. Seluruh proses ini bertujuan untuk menyederhanakan dan menstandarkan data teks sebelum masuk ke tahap klasifikasi.

2. Pelabelan Data

Pelabelan data dilakukan secara manual untuk memastikan akurasi dalam klasifikasi sentimen. Penelitian ini hanya menggunakan dua kategori sentimen, yaitu positif dan negative. Sentimen netral tidak disertakan karena sering kali sulit diinterpretasikan secara konsisten. Tweet dengan sentimen netral umumnya bersifat informatif atau ambigu, sehingga tidak secara tegas mencerminkan sikap pengguna terhadap layanan, baik berupa kepuasan maupun ketidakpuasan. Kondisi tersebut dapat menurunkan reliabilitas pelabelan. Oleh karena itu, pembatasan analisis pada dua kelas utama memungkinkan proses klasifikasi yang lebih efisien serta menghasilkan interpretasi hasil yang lebih reliabel.

Proses pelabelan ini dilakukan dengan menganalisis setiap komentar berdasarkan konteksnya, tanpa menggunakan metode otomatis, sehingga menghasilkan data yang lebih sesuai untuk analisis sentimen. Hasil dari pelabelan manual menunjukkan bahwa data terdiri dari 5319 sentimen negatif dan 2074 sentimen positif. Beberapa contoh tweet dapat dilihat pada Tabel 1.

Tabel 1. Data Contoh Tweet dan Label Sentimen.

Tweet	Sentimen
indihome jelek banget	negatif
indihome sampah	negatif
bahagia banget pakai indihome selalu lancar	positif
indihome lancar alhamdulillah	positif

3. Pembagian Data dan Ekstraksi Fitur

Data yang telah melalui tahapan preprocessing dan pelabelan dibagi menjadi dua bagian, yaitu 90% sebagai data latih dan 10% sebagai data uji. Pembagian ini dilakukan secara konsisten pada semua model. Dari total 7393 data tweet, sebanyak 6653 data digunakan sebagai data latih, dan 740 data sisanya digunakan sebagai data uji.

Data teks kemudian diubah menjadi fitur numerik menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* untuk model machine learning tradisional. TF-IDF digunakan untuk merepresentasikan seberapa penting suatu kata dalam sebuah dokumen relatif terhadap seluruh korpus, sehingga menghasilkan vektor fitur yang dapat digunakan dalam proses pelatihan model. Sementara itu, untuk model IndoBERT, proses tokenisasi dilakukan menggunakan tokenizer dari IndoBERT-base-p1. Tokenizer ini mengubah teks mentah menjadi representasi numerik melalui subword tokenization dengan metode *WordPiece*. Setiap teks dikonversi menjadi token sesuai dengan vocabulary IndoBERT, dengan penambahan token khusus [CLS] dan [SEP], lalu diubah menjadi ID numerik. Untuk menjaga keseragaman panjang input, teks yang lebih pendek akan diberi *padding* dan yang lebih panjang akan dipotong (*truncated*). *Attention mask* juga dihasilkan untuk membedakan token bermakna dari token *padding*.

4. Penanganan Ketidakseimbangan Data

Strategi penanganan ketidakseimbangan data disesuaikan dengan karakteristik masing-masing model. Pada model Naïve Bayes, yang berbasis pada representasi numerik TF-IDF, digunakan teknik *oversampling* berupa SMOTE yang mengatasi ketidakseimbangan dengan membuat sampel sintetis dari

kelas minoritas [20]. Untuk model SVM, diterapkan dua pendekatan, yaitu SMOTE dan pembobotan kelas melalui parameter *class_weight='balanced'*. Parameter *class_weight='balanced'* secara otomatis menghitung bobot masing-masing kelas berdasarkan kebalikan dari frekuensi kemunculannya dalam dataset, sehingga kelas dengan frekuensi yang lebih rendah mendapatkan bobot yang lebih tinggi [21]. Adapun pada model IndoBERT, ketidakseimbangan kelas ditangani dengan pendekatan serupa, yakni pembobotan kelas yang diterapkan pada fungsi loss selama proses pelatihan.

5. **Pemodelan dan Klasifikasi**

Terdapat tiga model yang digunakan dalam menganalisis sentimen, yaitu dua model machine learning tradisional yaitu Naïve Bayes dan SVM serta satu model berbasis transformer, yaitu IndoBERT. Untuk model Naive Bayes dan SVM, konfigurasi awal digunakan secara default. Sementara itu, model IndoBERT yang memiliki arsitektur lebih kompleks dan performa yang lebih sensitif terhadap perubahan *hyperparameter*, dilakukan proses tuning ringan guna menyesuaikan dengan karakteristik data. Proses ini tetap dijaga agar tidak terlalu agresif, demi menjaga keadilan dalam perbandingan antar model. Adapun hyperparameter pada masing-masing model, Naïve Bayes menggunakan alpha sebagai penghalusan (*smoothing*) untuk menghindari pembagian dengan nol dan menjaga stabilitas ketika menghadapi fitur yang jarang. SVM menggunakan C untuk mengatur margin dan kesalahan, kernel untuk memetakan data non-linear, dan gamma untuk mengontrol pengaruh tiap titik data. IndoBERT menggunakan *learning rate* untuk pembaruan bobot, *batch size* untuk jumlah data per pembaruan, *dropout rate* sebagai persentase neuron yang dinonaktifkan secara acak selama pelatihan untuk mencegah overfitting, dan *epochs* sebagai jumlah siklus pelatihan.

Tabel 2. Hyperparameter Model.

Model	Hyperparameter	Nilai
Naïve Bayes	alpha	1.0
	C	1.0
SVM	kernel	rbf
	gamma	scale
	Learning rate	1.62e-05
IndoBERT	Batch size	32
	Dropout rate	0.26
	Epochs	4

6. **Pelatihan dan Evaluasi Model**

Untuk mengukur ketangguhan dan kestabilan performa model, digunakan metode *stratified 10-fold cross validation* pada data latih. Model dilatih dan divalidasi sebanyak sepuluh kali pada fold berbeda, dan rata-rata akurasi dari seluruh fold digunakan sebagai indikator awal performa model. Setelah validasi, model akhir dilatih ulang dengan seluruh data latih, lalu dievaluasi pada data uji.

Evaluasi performa dilakukan menggunakan sejumlah metrik, yaitu akurasi, *precision*, *recall*, dan *F1-score*. Akurasi mengukur persentase data yang diprediksi dengan benar dari seluruh data, *precision* menunjukkan seberapa banyak data yang diprediksi positif memang benar-benar positif, *recall* menunjukkan seberapa banyak data positif yang berhasil dikenali dengan benar, sedangkan *F1-score* merupakan rata-rata harmonik dari *precision* dan *recall* [22]. Rumus dari masing-masing metrik dapat dituliskan sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN}, \tag{4}$$

$$Precision = \frac{TP}{TP+FP}, \tag{5}$$

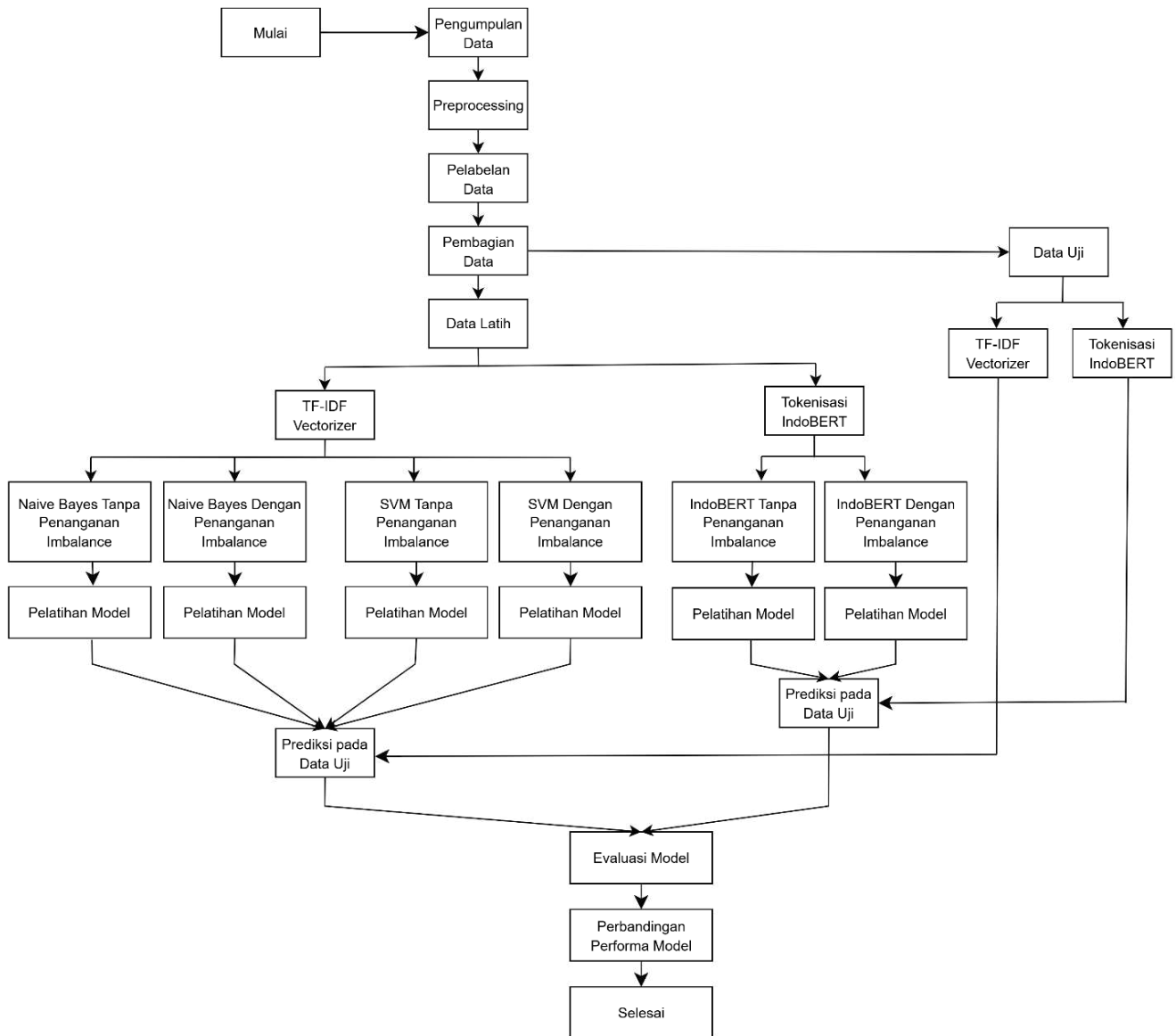
$$Recall = \frac{TP}{TP+FN}, \tag{6}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{7}$$

dengan TP (*True Positive*): jumlah prediksi positif yang benar, TN (*True Negative*): jumlah prediksi negatif yang benar, FP (*False Positive*): jumlah prediksi positif yang salah, dan FN (*False Negative*): jumlah prediksi negatif yang salah. Selain itu, digunakan juga F1-score macro yang dirumuskan sebagai rata-rata aritmatika dari F1-score tiap kelas [23]:

$$Macro - F1 = \frac{1}{k} \sum_{i=1}^k F1_i \tag{8}$$

Metrik *F1-score macro* dipilih pada tabel perbandingan antar model karena metrik ini memberikan penilaian rata-rata terhadap seluruh kelas secara seimbang, sehingga cocok digunakan dalam kondisi data yang tidak seimbang. Secara garis besar, pengolahan data yang dilakukan pada penelitian ini terangkum pada Gambar 3.



Gambar 1. Alur Penelitian Analisis Sentimen IndiHome.

Hasil dan Pembahasan

Berdasarkan data, hasil pelabelan pada data Twitter diperoleh distribusi kelas untuk sentimen positif dan negatif adalah sebagai berikut:

Tabel 3. Distribusi Awal Data Sentimen Twitter terhadap IndiHome.

Sentimen	Jumlah Data
Negatif	5319
Positif	2074
Total	7393

Berdasarkan Tabel 3, terdapat ketidakseimbangan distribusi kelas. Dari total 7393 tweet, kelas sentimen negatif mendominasi dengan jumlah 5319 data, diikuti oleh kelas positif sebanyak 2074. Ketidakseimbangan ini berpotensi memengaruhi kinerja model, khususnya dalam mengenali sentimen minoritas [24].

Evaluasi Model Tanpa Penanganan Ketidakseimbangan Data

Bagian ini menyajikan hasil evaluasi model klasifikasi sentimen tanpa penanganan terhadap ketidakseimbangan data. Tujuan dari analisis ini adalah untuk mengetahui performa awal dari masing-masing model terhadap distribusi data yang tidak seimbang. Untuk melihat stabilitas model, dilakukan 1 – 10 fold cross validation pada data latih untuk model Naïve Bayes, SVM, dan IndoBERT. Hasil akurasi tertera pada Tabel 4. Berdasarkan tabel ini, model Naïve Bayes memiliki akurasi yang stabil, berkisar antara 0.89 hingga 0.92, dengan rata-rata sebesar 0.90. Model SVM tampil konsisten dengan akurasi antara 0.92 hingga 0.95, menghasilkan rata-rata 0.94. Sementara itu, IndoBERT mencatatkan akurasi tertinggi di antara ketiganya, yaitu antara 0.94 hingga 0.97, dengan rata-rata 0.96. Hasil validasi ini menunjukkan bahwa, pada tahap pelatihan, IndoBERT menunjukkan potensi kinerja yang lebih unggul dibandingkan dua model tradisional lainnya. Namun, performa akhir model tetap perlu dikonfirmasi melalui evaluasi pada data uji terpisah.

Tabel 4. Hasil Akurasi Stratified 10-Fold Cross Validation.

Fold	Naïve Bayes	SVM	IndoBERT
1	0.90	0.92	0.96
2	0.89	0.95	0.96
3	0.90	0.93	0.96
4	0.89	0.93	0.94
5	0.89	0.93	0.96
6	0.92	0.94	0.97
7	0.92	0.94	0.97
8	0.91	0.94	0.97
9	0.90	0.94	0.97
10	0.89	0.94	0.95
Rata-rata	0.90	0.94	0.96

Setelah dilakukan evaluasi melalui cross-validation, model akhir diujikan pada data uji untuk mengukur performa generalisasi. Performa klasifikasi menggunakan model Naïve Bayes tertera pada Tabel 5. Pada model ini, terlihat bahwa kelas negatif menunjukkan performa yang sangat baik dengan nilai recall sebesar 0.99 dan precision sebesar 0.88, yang berarti model mampu mengenali hampir seluruh

data Negatif dengan cukup akurat. Di sisi lain, performa pada kelas positif menunjukkan ketidakseimbangan, dengan precision yang sangat tinggi 0.97 namun recall yang rendah 0.61. Hal ini mengindikasikan bahwa meskipun prediksi positif yang dihasilkan sangat tepat, model hanya mampu mengenali sebagian kecil dari seluruh data positif yang ada. Akurasi keseluruhan model mencapai 0.89, yang tergolong baik, tetapi adanya perbedaan performa yang mencolok antar kelas menjadi perhatian dalam evaluasi model ini.

Tabel 5. Performa Klasifikasi Naive Bayes.

Sentimen	precision	recall	F1-score
Negatif	0.88	0.99	0.93
Positif	0.97	0.61	0.74
Akurasi			0.89

Performa klasifikasi menggunakan model SVM tertera pada Tabel 6. Kinerja menggunakan model SVM lebih seimbang dibandingkan model Naive Bayes. Kelas negatif mencapai nilai precision sebesar 0.95 dan recall sebesar 0.98, menghasilkan F1-score yang sangat tinggi, yaitu 0.96. Hal ini menunjukkan bahwa model sangat andal dalam mengenali sentimen negatif tanpa banyak kesalahan. Sementara itu, kelas positif juga memperoleh precision yang tinggi sebesar 0.93, namun recall-nya sedikit lebih rendah, yaitu 0.84, menghasilkan F1-score sebesar 0.88. Secara keseluruhan, akurasi model SVM mencapai 0.94, menunjukkan peningkatan signifikan dibandingkan dengan Naive Bayes. Hasil ini sejalan dengan temuan dalam [25][26], bahwa SVM lebih mampu menangani kompleksitas data sentimen berbasis media sosial dibandingkan Naive Bayes.

Tabel 6. Performa Klasifikasi SVM.

Sentimen	Precision	Recall	F1-score
Negatif	0.95	0.98	0.96
Positif	0.93	0.84	0.88
Akurasi			0.94

Performa klasifikasi menggunakan model IndoBERT tertera pada Tabel 7. Model IndoBERT memberikan performa tertinggi di antara ketiga model yang diuji. Pada kelas negatif, precision dan recall mencapai 0.97, menghasilkan F1-score yang sangat tinggi sebesar 0.97. Artinya, model sangat andal dalam mendeteksi dan mengklasifikasikan data negatif secara akurat dan konsisten. Sementara itu, kelas positif memperoleh precision dan recall sebesar 0.92, dengan F1-score yang sama 0.92, yang juga tergolong sangat baik dan menunjukkan bahwa model mampu mengenali sebagian besar data positif dengan tingkat kesalahan yang relatif rendah. Akurasi total sebesar 0.96 mengindikasikan bahwa IndoBERT adalah model dengan performa terbaik dalam klasifikasi sentimen pada dataset ini, dengan keseimbangan antara precision dan recall yang tinggi untuk kedua kelas.

Tabel 7. Performa Klasifikasi IndoBERT.

Sentimen	Precision	Recall	F1-score
Negatif	0.97	0.97	0.97
Positif	0.92	0.92	0.92
Akurasi			0.96

Untuk memperkuat klaim keunggulan model, dilakukan pengujian statistik terhadap hasil akurasi dari ketiga model berdasarkan hasil Stratified 10-Fold Cross Validation (Tabel 4). Berdasarkan hasil perhitungan confidence interval (95%), model Naive Bayes memperoleh rata-rata akurasi sebesar 0.9010

± 0.0086 , SVM sebesar 0.9360 ± 0.0060 , dan IndoBERT sebesar 0.9610 ± 0.0071 . Nilai ini menunjukkan bahwa IndoBERT memiliki rata-rata akurasi tertinggi dengan rentang ketidakpastian yang relatif kecil, menandakan performa yang stabil di berbagai fold.

Selanjutnya, untuk memastikan bahwa perbedaan performa antar model bersifat signifikan secara statistik, dilakukan uji Wilcoxon Signed-Rank Test terhadap hasil akurasi tiap fold. Hasil pengujian menunjukkan bahwa perbedaan antara setiap pasangan model signifikan dengan nilai p-value sebesar 0.002 ($p < 0.05$), baik antara Naïve Bayes dan SVM, Naïve Bayes dan IndoBERT, maupun SVM dan IndoBERT. Dengan demikian, dapat disimpulkan bahwa IndoBERT secara signifikan mengungguli Naïve Bayes dan SVM dalam performa klasifikasi sentimen pada dataset ini.

Temuan dalam penelitian ini yang menunjukkan keunggulan IndoBERT dibandingkan Naïve Bayes dan SVM sejalan dengan studi sebelumnya. Misalnya, [27] menyatakan bahwa dalam analisis sentimen terhadap tweet terkait bencana alam, IndoBERT mencapai akurasi sebesar 91%, jauh lebih tinggi dibandingkan dengan Naïve Bayes yang hanya mencapai 74%. Selain itu, [28] juga menunjukkan bahwa sebagian besar model *deep learning*, termasuk IndoBERT, mampu mengungguli baseline SVM. Hasil-hasil tersebut memperkuat temuan bahwa model berbasis transformer seperti IndoBERT memiliki kapabilitas yang lebih tinggi dalam memahami konteks teks dan menangani kompleksitas bahasa alami dibandingkan model pembelajaran mesin tradisional.

Evaluasi Model Dengan Penanganan Ketidakseimbangan Data

Pada bagian ini akan disajikan hasil evaluasi model klasifikasi sentimen setelah dilakukan penanganan terhadap ketidakseimbangan data. Tujuan dari evaluasi ini adalah untuk menilai sejauh mana teknik penanganan ketidakseimbangan mampu meningkatkan performa masing-masing model, baik dari segi kestabilan saat pelatihan (cross-validation) maupun akurasi pada data uji. Tabel 8 menunjukkan distribusi kelas pada data pelatihan sebelum dan sesudah dilakukan oversampling menggunakan SMOTE.

Tabel 8. Perbandingan Distribusi Sentimen Sebelum dan Sesudah SMOTE.

Sentimen	Jumlah	
	Tanpa SMOTE	Dengan SMOTE
Negatif	4769	4769
Positif	1884	4769

Setelah dilakukan penanganan data, hasil akurasi 1 – 10 fold cross validation untuk model Naïve Bayes, SVM, dan IndoBERT ditampilkan pada Tabel 9. Berdasarkan tabel ini, model Naïve Bayes mengalami peningkatan rata-rata akurasi dari 0.90 menjadi 0.92 setelah diterapkannya SMOTE. Hal ini menunjukkan bahwa penyeimbangan data melalui oversampling memberikan dampak positif pada model ini. Pada model SVM menunjukkan rata-rata akurasi yang sama antara pendekatan SMOTE dan pembobotan kelas, yaitu sebesar 0.94. Namun pendekatan menggunakan pembobotan kelas menghasilkan sedikit peningkatan pada beberapa fold dan cenderung mempertahankan akurasi yang lebih tinggi. Sementara itu, model IndoBERT menunjukkan hasil yang konsisten pada setiap fold, dengan rata-rata akurasi sebesar 0.96 baik sebelum maupun sesudah penerapan pembobotan kelas pada fungsi loss, sehingga dapat disimpulkan bahwa pembobotan kelas tidak memberikan pengaruh signifikan terhadap performa model ini.

Tabel 9. Hasil Akurasi Stratified 1-10 Fold Cross Validation dengan Penanganan Data Tidak Seimbang.

Fold	Naïve Bayes (SMOTE)	SVM (SMOTE)	SVM (Pembobotan Kelas)	IndoBERT (Pembobotan Kelas)
1	0.92	0.92	0.93	0.95

2	0.94	0.94	0.95	0.95
3	0.91	0.93	0.94	0.96
4	0.93	0.94	0.94	0.95
5	0.92	0.93	0.93	0.96
6	0.93	0.94	0.95	0.97
7	0.92	0.93	0.94	0.97
8	0.92	0.95	0.94	0.96
9	0.93	0.94	0.95	0.97
10	0.90	0.94	0.93	0.95
Rata-rata	0.92	0.94	0.94	0.96

Ketiga model yang dihasilkan pada data latih kemudian diujikan pada data uji dan performa klasifikasi tertera pada Tabel 10 - 13. Performa klasifikasi tertera menggunakan Naive Bayes (SMOTE) disajikan pada Tabel 10. Model ini mengalami peningkatan performa yang cukup signifikan setelah diterapkan SMOTE. Recall pada kelas positif meningkat dari 0.61 menjadi 0.88, yang menandakan bahwa model kini mampu mengenali lebih banyak data positif. F1-score kelas positif juga meningkat dari 0.74 menjadi 0.85. Akurasi keseluruhan model naik menjadi 0.92.

Tabel 10. Performa Klasifikasi Naive Bayes (SMOTE).

Sentimen	Precision	Recall	F1-score
Negatif	0.96	0.93	0.95
Positif	0.82	0.88	0.85
Akurasi			0.92

Performa klasifikasi tertera menggunakan SVM (SMOTE) disajikan pada Tabel 11. Pada model SVM, SMOTE hanya memberikan dampak yang sangat kecil, dengan sedikit peningkatan pada F1-score kelas positif dari 0.88 menjadi 0.89, namun tidak mempengaruhi akurasi secara keseluruhan. Hal ini menunjukkan bahwa SMOTE hanya memberi dampak kecil pada keseimbangan klasifikasi SVM. Selain itu, SVM diuji menggunakan pembobotan kelas otomatis melalui parameter `class_weight='balanced'`. Penerapan pembobotan kelas pada SVM meningkatkan akurasi menjadi 0.95, dengan F1-score kelas positif naik menjadi 0.90 dan kelas negatif menjadi 0.97 (lihat Tabel 12). Hal ini menunjukkan bahwa pembobotan kelas lebih efektif daripada SMOTE dalam meningkatkan performa SVM pada data tidak seimbang.

Tabel 11. Performa Klasifikasi SVM (SMOTE).

Sentimen	Precision	Recall	F1-score
Negatif	0.95	0.97	0.96
Positif	0.92	0.85	0.89
Akurasi			0.94

Tabel 12. Performa Klasifikasi SVM (pembobotan kelas).

Sentimen	Precision	Recall	F1-score
Negatif	0.96	0.97	0.97
Positif	0.92	0.89	0.90
Akurasi			0.95

Untuk model IndoBERT (class wighting), performa ditunjukkan pada Tabel 13. Berdasarkan Tabel 13, klasifikasi menggunakan model IndoBERT (class wighting) menunjukkan performa terbaik, dengan keseimbangan precision dan recall yang tinggi pada kedua kelas. F1-score untuk kelas Positif naik dari

0.92 (tanpa pembobotan) menjadi 0.93. Meskipun kenaikannya relatif kecil, hasil ini menunjukkan bahwa pembobotan kelas tetap memberikan dampak positif. Penemuan ini konsisten dengan studi sebelumnya seperti yang dilakukan oleh [29] yang menerapkan teknik pembobotan kelas untuk penanganan ketidakseimbangan data dan menyatakan bahwa pendekatan tersebut meningkatkan F1-score dengan mengurangi underprediction terhadap kelas minoritas.

Tabel 13. Performa Klasifikasi IndoBERT (pembobotan kelas).

Sentimen	Precision	Recall	F1-score
Negatif	0.97	0.98	0.98
Positif	0.94	0.92	0.93
Akurasi			0.96

Dengan demikian, dapat dilihat bahwa teknik penanganan ketidakseimbangan data memberikan dampak yang bervariasi terhadap performa model. Naïve Bayes menunjukkan peningkatan signifikan setelah penerapan SMOTE, sedangkan SMOTE dan pembobotan kelas memberikan sedikit peningkatan pada SVM. Model IndoBERT relatif stabil dan menunjukkan performa tinggi bahkan tanpa penanganan tambahan. Penerapan pembobotan kelas memberikan peningkatan yang sangat kecil, namun secara keseluruhan tidak cukup signifikan untuk dianggap meningkatkan performa model secara substansial. Ringkasan performa akhir model pada data uji ditampilkan pada Tabel 14.

Tabel 14. Perbandingan Performa Seluruh Model pada Data Uji.

Model	Penanganan	Akurasi	F1-score Macro
Naïve Bayes	Tidak Ada	0.89	0.84
	SMOTE	0.92	0.90
SVM	Tidak Ada	0.94	0.92
	SMOTE	0.94	0.92
	Pembobotan kelas	0.95	0.93
IndoBERT	Tidak Ada	0.96	0.95
	Pembobotan kelas	0.96	0.95

Berdasarkan Tabel 14, peningkatan performa yang paling mencolok terlihat pada model Naïve Bayes setelah penerapan SMOTE, dengan kenaikan pada akurasi dari 0.89 menjadi 0.92 dan F1-score macro dari 0.84 menjadi 0.90. Hal ini menunjukkan bahwa SMOTE efektif meningkatkan performa model, terutama dalam menangani ketidakseimbangan data dan mengenali kelas minoritas. Hasil ini diperkuat oleh temuan dalam penelitian oleh [30] melaporkan bahwa akurasi Naive Bayes meningkat dari 78.18% menjadi 81.90% setelah penerapan SMOTE.

Pada model SVM, SMOTE tidak memberikan perubahan yang berarti, sementara pembobotan kelas justru meningkatkan akurasi dan F1-score macro, menjadikannya strategi yang lebih sesuai untuk model tersebut. Temuan ini sejalan dengan penelitian oleh [31], yang menyatakan bahwa pembobotan kelas secara seimbang (seperti `class_weight = 'balanced'` untuk SVM dan Regresi Logistik, dan bobot serupa untuk model XGBoost dan Random Forest) mampu meningkatkan kinerja model terhadap kelas minoritas tanpa menyebabkan overfitting, berbeda dengan metode oversampling seperti SMOTE yang justru menyebabkan overfitting pada beberapa model. IndoBERT tetap menjadi model dengan performa tertinggi di antara seluruh model yang diuji. Tanpa penanganan khusus pun model ini sudah mencapai akurasi 0.96. Namun penerapan pembobotan kelas pada model IndoBERT tidak memberikan peningkatan performa yang signifikan. Akurasi model tetap berada pada angka 0.96, dan F1-score macro juga tetap berada di angka 0.95.

Meskipun data tidak seimbang, penerapan penanganan ketidakseimbangan data pada SVM dan IndoBERT tidak memberikan peningkatan performa yang signifikan. Hal ini diduga karena performa awal model sudah tergolong tinggi, sehingga ruang perbaikannya terbatas. Selain itu, jumlah data kelas minoritas secara absolut masih cukup besar dan representatif, serta model memiliki kemampuan representasi fitur yang kuat yaitu TF-IDF pada SVM dan contextual embedding pada IndoBERT. IndoBERT sendiri dikenal cukup robust terhadap ketidakseimbangan data karena dilatih secara luas pada beragam korpus. Di sisi lain, Naïve Bayes menunjukkan peningkatan setelah penerapan SMOTE karena meskipun performa awal Naïve Bayes cukup tinggi secara akurasi, ketidakseimbangan performa antar kelas sangat terlihat, sehingga masih terdapat ruang perbaikannya. Selain itu, algoritma Naïve Bayes sangat bergantung pada estimasi probabilitas fitur terhadap kelas, sehingga penambahan data sintesis membantu mengurangi bias pada kelas minoritas.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pada penelitian ini SMOTE paling efektif diterapkan pada Naïve Bayes untuk meningkatkan sensitivitas terhadap kelas minoritas. Sementara itu, pembobotan kelas lebih sesuai digunakan pada SVM. IndoBERT, sebagai model berbasis pretrained language model, telah menunjukkan performa tinggi tanpa penanganan khusus, dan hanya mengalami peningkatan kecil setelah pembobotan kelas. Temuan ini menegaskan bahwa efektivitas teknik penanganan ketidakseimbangan data sangat bergantung pada jenis model yang digunakan serta konteks data yang dianalisis.

Word Cloud dan Tren Sentimen

Untuk melengkapi analisis hasil model, dilakukan proses prediksi terhadap seluruh data yang tersedia. Langkah ini tidak bertujuan untuk mengevaluasi performa model seperti akurasi atau metrik lainnya, melainkan untuk memahami kecenderungan model dalam memetakan distribusi sentimen. Informasi ini dimanfaatkan lebih lanjut dalam visualisasi word cloud dan analisis tren sentimen berdasarkan distribusi sentimen yang diprediksi oleh model terbaik. Distribusi hasil prediksi model terbaik, yaitu IndoBERT dengan pembobotan kelas dibandingkan dengan distribusi data asli ditampilkan pada Tabel 15.

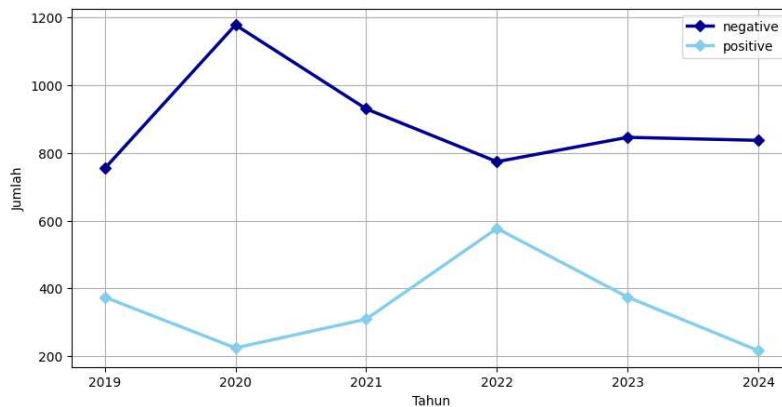
Tabel 15. Distribusi Sentimen Asli vs Prediksi IndoBERT (pembobotan kelas).

Sentimen	Distribusi Asli	Prediksi IndoBERT
Negatif	5319	5325
Positif	2074	2068

Dari Tabel 15, terlihat bahwa distribusi prediksi sangat mendekati distribusi sentimen asli. Perbedaan jumlah antara distribusi asli dan hasil prediksi hanya berselisih 6 data pada masing-masing kelas, yang menunjukkan bahwa model memiliki kecenderungan prediksi yang sangat seimbang terhadap dua kelas sentimen. Kelas negatif sedikit lebih banyak diprediksi daripada jumlah aslinya, namun selisih tersebut relatif kecil dan tidak menunjukkan adanya bias signifikan terhadap salah satu kelas. Keseimbangan prediksi ini menunjukkan bahwa IndoBERT mampu menangkap pola distribusi sentimen dalam data secara menyeluruh. Oleh karena itu, hasil prediksi ini dianggap representatif dan dapat digunakan sebagai dasar untuk analisis lanjut dalam visualisasi word cloud dan analisis tren sentimen.

Untuk memberikan gambaran visual mengenai sentimen yang terkandung dalam data, berikut ditampilkan word cloud yang menggambarkan kata-kata yang paling sering muncul dalam sentimen positif dan negatif.

Beberapa tweet yang mewakili word cloud sentimen negatif ini antara lain, “Kacau lambat IndiHome anjing”, “IndiHome eror terus sih, bikin kesal lagi banyak deadline gitu”, “Anjing ngeselin memang IndiHome, tiba-tiba WiFi mati”, dan “Tagihan Telkom IndiHome mendadak naik”. Kutipan-kutipan ini memperkuat hasil visualisasi word cloud tersebut. Temuan ini menjadi bahan evaluasi penting bagi IndiHome untuk memprioritaskan perbaikan pada aspek kecepatan internet, kestabilan koneksi, dan sistem penagihan. Selain itu, peningkatan kualitas layanan pelanggan dan mekanisme penanganan keluhan yang responsif juga diperlukan, mengingat tingginya ekspresi frustrasi dalam tweet pelanggan.



Gambar 4. Tren Sentimen IndiHome per Tahun.

Grafik menunjukkan bahwa sepanjang periode 2019–2024, sentimen negatif secara konsisten mendominasi dibandingkan sentimen positif. Puncak keluhan terjadi pada tahun 2020, di mana jumlah tweet negatif mencapai titik tertinggi. Hal ini kemungkinan berkaitan dengan peningkatan penggunaan internet selama masa pandemi COVID-19, hal ini sejalan dengan temuan penelitian bahwa pandemi tersebut menyebabkan peningkatan lalu lintas internet sebesar 15-20% hanya dalam waktu satu minggu [33].

Setelah 2020, jumlah sentimen negatif menurun pada 2021 dan relatif stabil hingga 2024, meskipun tetap berada pada angka yang tinggi. Sementara itu, sentimen positif mengalami fluktuasi, dengan kenaikan tertinggi pada tahun 2022, yang kemungkinan mengindikasikan adanya perbaikan layanan atau strategi promosi yang berhasil pada tahun tersebut. Namun, tren positif ini tidak bertahan lama, karena terjadi penurunan kembali pada 2023 dan terus menurun hingga 2024.

Secara keseluruhan, tren ini menunjukkan bahwa meskipun terdapat upaya perbaikan, persepsi negatif terhadap layanan IndiHome masih cukup dominan. Tren ini menunjukkan perlunya peningkatan kualitas layanan secara berkelanjutan dan respons yang lebih cepat terhadap keluhan pelanggan. IndiHome juga dapat melakukan evaluasi mendalam terhadap strategi yang diterapkan pada tahun 2022, yang tampaknya berhasil meningkatkan sentimen positif, untuk diterapkan kembali secara lebih konsisten di tahun-tahun berikutnya. Analisis lebih lanjut terhadap penyebab naik turunnya persepsi pengguna dapat menjadi dasar untuk pengambilan keputusan strategis berbasis data.

Kesimpulan

Penelitian ini membandingkan performa tiga model dalam analisis sentimen Twitter terkait layanan IndiHome, yaitu Naïve Bayes, Support Vector Machine (SVM), dan IndoBERT. Hasil menunjukkan bahwa IndoBERT memberikan performa terbaik dengan akurasi 0,96 dan F1-score makro 0,95, menegaskan keunggulan model berbasis transformer dalam memahami konteks bahasa Indonesia. SVM menunjukkan kinerja lebih baik daripada Naïve Bayes, terutama dengan penerapan pembobotan kelas yang meningkatkan akurasi hingga 0,95. Sementara itu, Naïve Bayes memperoleh peningkatan signifikan

setelah penerapan SMOTE, dari akurasi 0,89 menjadi 0,92 dan F1-score makro dari 0,84 menjadi 0,90. Temuan ini menunjukkan bahwa strategi penanganan data tidak seimbang lebih berdampak pada model tradisional, sedangkan IndoBERT relatif stabil bahkan tanpa penanganan tambahan. Selain evaluasi kuantitatif, analisis word cloud dan tren sentimen mengonfirmasi bahwa opini negatif masih mendominasi persepsi publik terhadap IndiHome, terutama terkait kecepatan dan kestabilan layanan, dengan puncak keluhan terjadi pada masa pandemi tahun 2020. Kontribusi utama penelitian ini adalah memberikan bukti empiris bahwa IndoBERT lebih unggul dalam klasifikasi sentimen berbahasa Indonesia dibandingkan model tradisional, sekaligus menunjukkan relevansi teknik penyeimbangan data untuk meningkatkan kinerja algoritma sederhana seperti Naïve Bayes. Namun, penelitian ini memiliki keterbatasan, antara lain data hanya bersumber dari Twitter dan fokus pada dua kelas sentimen (positif dan negatif). Penelitian selanjutnya dapat memperluas analisis ke platform media sosial lain, menambahkan kategori sentimen netral, serta mengeksplorasi model transformer lain seperti IndoRoBERTa atau IndoGPT untuk melihat perbandingan performa yang lebih luas.

Referensi

- [1] Simon Kemp, "Digital 2024: Indonesia," 2024. [Online]. Available: <https://datareportal.com/reports/digital-2024-indonesia>
- [2] APJII, "Profil Internet Indonesia 2022," Jakarta. [Online]. Available: <https://survei.apjii.or.id/survei>
- [3] F. Syah, H. Fajrin, A. N. Afif, M. R. Saeputra, D. Mirranty, and D. D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, pp. 53–58, Jan. 2023, doi: 10.35870/jtik.v7i1.686.
- [4] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *Eduatic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, Nov. 2020, doi: 10.21107/edutic.v7i1.8779.
- [5] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, Aug. 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [6] D. Darwis, N. Siskawati, and Z. Abidin, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, Feb. 2021, doi: 10.33365/jtk.v15i1.744.
- [7] Y. MZ, J. Bororing Edwin, S. Rahayu, and J. F. Andhika, "Analisis Sentimen Terhadap Layanan Tokopedia Berdasarkan Twitter dengan Metode Klasifikasi Support Vector Machine," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 12, no. 1, Jan. 2023, doi: 10.30591/smartcomp.v12i1.4591.
- [8] K. Munawaroh and A. Alamsyah, "Performance Comparison of SVM, Naïve Bayes, and KNN Algorithms for Analysis of Public Opinion Sentiment Against COVID-19 Vaccination on Twitter," *J. Adv. Inf. Syst. Technol.*, vol. 4, no. 2, pp. 113–125, Mar. 2023, doi: 10.15294/jaist.v4i2.59493.
- [9] N. P. A. P. S. Putri, D. S. Angreni, and I. W. Sudarsana, "A Study on Sentiment Analysis of Public Response to The New Fuel Price Policy In 2022: A Support Vector Machine Approach," *Inpr. Indones. J. Pure Appl. Math.*, vol. 7, no. 1, pp. 88–100, May 2025, doi: 10.15408/inprime.v7i1.42717.
- [10] F. Hashfi, D. Sugiarto, and I. Mardianto, "Sentiment Analysis of An Internet Provider Company Based on Twitter Using Support Vector Machine and Naïve Bayes Method," *Ultim. J. Tek. Inform.*, vol. 14, no. 1, pp. 1–6, 2022, doi: 10.31937/ti.v14i1.2384.
- [11] D. Pratama and S. Akbar, "Analysis of Public Opinion on Public Transportation in Bandung and Jakarta in Twitter using Indonesian Bidirectional Encoder Representations from Transformer," in *2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, BALI, Indonesia: IEEE, 2023, pp. 179–183. doi: 10.1109/IAICT59002.2023.10205608.
- [12] A. Rahmawati, A. Alamsyah, and A. Romadhony, "Hoax News Detection Analysis using IndoBERT Deep Learning Methodology," *2022 10th Int. Conf. Inf. Commun. Technol. ICoICT 2022*, no. April, pp. 368–373, 2022, doi: 10.1109/ICoICT55009.2022.9914902.
- [13] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks," *Appl. Sci.*, vol. 13, no. 6, 2023, doi: 10.3390/app13064006.
- [14] B. Phatcharathada and P. Srisuradetchai, "Randomized Feature and Bootstrapped Naive Bayes Classification," *Appl. Syst. Innov.*, vol. 8, no. 4, p. 94, Jul. 2025, doi: 10.3390/asi8040094.
- [15] Y. A. Singgalen, "Comparative analysis of decision tree and support vector machine algorithm in sentiment classification for birds

- of paradise content,” *Int. J. Basic Appl. Sci.*, vol. 12, no. 3, pp. 100–109, Dec. 2023, doi: 10.35335/ijobas.v12i3.298.
- [16] K. K. Sampath and M. Supriya, “Transformer Based Sentiment Analysis on Code Mixed Data,” *Procedia Comput. Sci.*, vol. 233, no. 2023, pp. 682–691, 2024, doi: 10.1016/j.procs.2024.03.257.
- [17] E. Y. Zhang, A. D. Cheok, Z. Pan, J. Cai, and Y. Yan, “From Turing to Transformers: A Comprehensive Review and Tutorial on the Evolution and Applications of Generative Transformer Models,” *Sci*, vol. 5, no. 4, p. 46, Dec. 2023, doi: 10.3390/sci5040046.
- [18] B. Wilie *et al.*, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” 2020, [Online]. Available: <http://arxiv.org/abs/2009.05387>
- [19] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.
- [20] I. G. B. A. Budaya and I. K. P. Suniantara, “Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1077–1086, Jul. 2024, doi: 10.57152/malcom.v4i3.1459.
- [21] M. N. Razali, N. Arbaiy, P.-C. Lin, and S. Ismail, “Optimizing Multiclass Classification Using Convolutional Neural Networks with Class Weights and Early Stopping for Imbalanced Datasets,” *Electronics*, vol. 14, no. 4, p. 705, Feb. 2025, doi: 10.3390/electronics14040705.
- [22] O. Rainio, J. Teuvo, and R. Klén, “Evaluation metrics and statistical tests for machine learning,” *Sci. Rep.*, vol. 14, no. 1, p. 6086, Mar. 2024, doi: 10.1038/s41598-024-56706-x.
- [23] K. Takahashi, K. Yamamoto, A. Kuchiba, and T. Koyama, “Confidence interval for micro-averaged F1 and macro-averaged F1 scores,” *Appl. Intell.*, vol. 52, no. 5, pp. 4961–4972, Mar. 2022, doi: 10.1007/s10489-021-02635-5.
- [24] V. Ganganwar and R. Rajalakshmi, “Employing synthetic data for addressing the class imbalance in aspect-based sentiment classification,” *J. Inf. Telecommun.*, vol. 8, no. 2, pp. 167–188, Apr. 2024, doi: 10.1080/24751839.2023.2270824.
- [25] F. Amandasari and D. Damayanti, “Perbandingan Kinerja Support Vector Machine dan Naive Bayes dalam Klasifikasi Sentimen Twitter Terhadap Pelayanan BPJS,” *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 3, pp. 645–653, 2025, doi: 10.52436/1.jpti.680.
- [26] V. Agresia and R. R. Suryono, “Comparison of SVM, Naive Bayes, and Logistic Regression Algorithms for Sentiment Analysis of Fraud and Bots in Purchasing Concert Ticket,” *INOVTEK Polbeng - Seri Inform.*, vol. 10, no. 2, pp. 591–602, Jul. 2025, doi: 10.35314/npyfhd47.
- [27] S. M. Anugerah, R. Wijaya, and M. A. Bijaksana, “Sentimen Analysis Social Media for Disaster using Naive Bayes and IndoBERT,” *INTEK J. Penelit.*, vol. 11, no. 1, pp. 51–58, Apr. 2024, doi: 10.31963/intek.v11i1.4771.
- [28] W. O. Vihikan and I. N. P. Trisna, “Indonesian Health Question Multi-Class Classification Based on Deep Learning,” *J. Inf. Syst. Informatics*, vol. 6, no. 3, pp. 1931–1944, Sep. 2024, doi: 10.51519/journalisi.v6i3.838.
- [29] S. Rohan *et al.*, “BD at BEA 2025 Shared Task: MPNet Ensembles for Pedagogical Mistake Identification and Localization in AI Tutor Responses,” 2025, [Online]. Available: <http://arxiv.org/abs/2506.01817>
- [30] 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,” *Technol. Sci.*, vol. 7, no. 1, pp. 169–178, 2025, doi: <https://doi.org/10.47065/bits.v7i1.7261>.
- [31] N. Hussain *et al.*, “Multi-Level Depression Severity Detection with Deep Transformers and Enhanced Machine Learning Techniques,” *AI*, vol. 6, no. 7, p. 157, Jul. 2025, doi: 10.3390/ai6070157.
- [32] S. Widagdo, Y. I. Handayani, A. H. Prastyowati, L. Rachmawati, M. Dimiyati, and S. Amalia, “The effect of reliability and empathy on customer satisfaction: A survey of PT Telkom Indonesia’s IndiHome customers,” *Hum. Syst. Manag.*, vol. 43, no. 2, pp. 181–194, Mar. 2024, doi: 10.3233/HSM-230003.
- [33] A. Feldmann *et al.*, “The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic,” *Proc. ACM SIGCOMM Internet Meas. Conf. IMC*, pp. 1–18, 2020, doi: 10.1145/3419394.3423658.